

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 12, 2002, 03:22:00 ; Search time 2485 Seconds

(without alignments) 1244.804 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAMALPSILRLGAQETE.....LSPGNQLYHLIQQNPHYNSP 191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV-xtlp
-O/cgpn2.1/USPRO.spool/US09462625/runat_07112002.101526.24854/app.query.fasta.1.327
-DB-EST -QFMT-fastap -SUFFIX-rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS-bits -START-1 -END-1 -MATRIX-Biosum62 -TRANS-human4.0.cdi -LIST-45
-DOCALLIGN=200 -THR.SCORE-pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE-LOCAL
-OUTFMT-pto -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-US09462625.ecgn.1.1.2024 @runat.07112002.101526.24854 -NCPU=6 -ICPU=3
-NO.XLPXY -NO.MAP -LARGEROXY -NEG.SCORES=0 -WAIT -LONGLOS -DEV.TIMEOUT=120
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:
2: em_esthum:
3: em_estin:
4: em_estmu:
5: em_estov:
6: em_estpl:
7: em_estro:
8: em_hlc:
9: gb_est1:
10: gb_est2:
11: gb_hlc:
12: gb_est3:
13: gb_est4:
14: gb_est5:
15: em_estfun:
16: em_estom:
17: gb_gss:
18: em_gss_hum:
19: em_gss_iny:
20: em_gss_pln:
21: em_gss_vrt:
22: em_gss_fun:
23: em_gss_mam:
24: em_gss_mus:
25: em_gss_other:
26: em_gss_pro:
27: em_gss_rtd:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	791	75.3	575	10	AM467554
C 2	692	65.8	520	10	AM076051
C 3	692	65.8	589	12	BG174272
C 4	692	65.8	637	13	B1149595
C 5	692	65.8	676	13	B1508838
C 6	692	65.8	682	11	AK008335
C 7	692	65.8	686	14	BQ951956
C 8	692	65.8	688	12	BG244455
C 9	692	65.8	703	13	B1453419
C 10	690	65.7	650	13	B1155774
C 11	687	65.4	610	13	B1154844
C 12	678	64.5	600	12	BG871384
C 13	677	64.4	627	9	AA228200
C 14	675	64.2	645	13	BG975104
C 15	671	63.8	614	13	BE199698
C 16	669	63.7	842	13	B1409815
C 17	668	63.6	1045	12	BE913706
C 18	665	63.3	706	12	BF163190
C 19	658	62.6	619	12	BF302505
C 20	654	62.2	555	12	BF076851
C 21	653.5	60.3	601	9	AA238752
C 22	632	60.1	580	13	B1556032
C 23	620	59.0	575	9	AI507116
C 24	616	58.6	591	9	AA238564
C 25	602	57.3	513	14	BQ564677
C 26	586	55.8	528	9	AA734993
C 27	575.5	54.8	534	9	AA734805
C 28	567	53.9	547	9	AA689693
C 29	553	52.6	548	9	AA597240
C 30	529	50.3	548	9	AI585767
C 31	523	48.8	518	10	BE135520
C 32	521	49.6	438	13	BI849729
C 33	510	48.5	384	10	BE654595
C 34	503	47.9	409	12	BG187105
C 35	489	46.5	564	9	AA99942
C 36	481	45.8	510	12	BF152060
C 37	477	45.4	665	13	BG969181
C 38	463	44.1	384	9	AA689633
C 39	462	44.0	573	9	AA543570
C 40	411	39.1	451	12	BF151994
C 41	405.5	38.6	649	9	AI108979
C 42	395.5	37.6	625	10	AM941788
C 43	394.5	37.5	604	13	B1504243
C 44	385.5	36.7	649	13	B1504245
C 45	382	36.3	472	13	BI291182

ALIGNMENTS

RESULT 1
AM467554 LOCUS
DEFINITION he21d02.x1 NCI CGAP CML1 Homo sapiens cDNA clone IMAGE:2919651 3'
similar to TR:075594 OT5599 PEPTIDOGLYCAN RECOGNITION PROTEIN
PRECUSOR. ; mRNA sequence.
ACCESSION AM467554.1 GI:7037660
VERSION AM467554
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS NCI-Cgap <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
 Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.llnl.gov/dbp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 359.

FEATURES

source
 1. 575
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2919651"
 /clone_1lb="NCI CGAP CML1"
 /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
 rearrangement positive, includes both chronic phase and
 myeloid blast crisis"
 /lab_host="DH10B"
 /note="Organ: whole blood; Vector: PCMV-SPORT6; Site: 1;
 Salt: Site 2: NotI; Cloned unidirectionally. Primer:
 Oligo dT. Library constructed by Life Technologies."
 BASE COUNT 110 a 147 c 178 g 135 t 5 others
 ORIGIN

Alignment Scores:

Pred. No.: 6.75e-77 Length: 575
 Score: 791.00 Matches: 147
 Percent Similarity: 93.67% Conservative: 1
 Best Local Similarity: 93.04% Mismatches: 10
 Query Match: 75.26% Indels: 1
 DB: 10 Gaps: 0

US-09-462-625-4 (1-191) x AW467554 (1-575)

QY 34 TTPYSAALAEUASERGLUCYSAAGLNHLSERLEUPROLEUAQTYRVAVAL 53
 |||||
 DB 575 TGAAGGCCCTGGCATCAGATGCGCCACAGCTGGAGCTGCTTACGCTATGTGGTG 516
 QY 54 ValSerHisThrAlaGlySerSerCysAsnThrProAlaSerCysGlnGlnAlaArg 73
 |||||
 DB 515 GATATGCACCGCGGCGGAGCAGTGTGAACACCCCGCTTG-TGCAGCAAGCAGCCCGG 457
 QY 74 AsnValGlnHisThrMetLeuThrLeuGlyTTPCysAspValGlyTyrAsnPhelau 93
 |||||
 DB 456 ATGTGCAGACTACACATGAAAGACATGGGCTGTGCGACCTGGGCTCAACTTCCTG 397
 QY 94 IleGlyIuAspGlyLeuValTyrGlnGlyArgGlyTyrAsnPherHrgIuAlaHisSer 113
 |||||
 DB 396 ATTGAGAGAGAGCGGCTGATACGAGGCGCGTGGTGAACCTTCAGGGTGGCCCTCA 337
 QY 114 GlyHisLeuThrAsnProMetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArg 133
 |||||
 DB 336 GGTCACTTATGGAACCCATGTCCATTTGCATCAGCTCATGGCAANTACATGATCGG 277
 QY 134 ValProThrProGlnAlaIleArgAlaIleGlnGlyLeuLeuAlaGlyValAlaGln 153
 |||||
 DB 276 GTGCCACACACCCAGGCCATCCGGGACGCCAGGGTTTATGGCCGTGGGTGGCTCAG 217
 QY 154 GlyAlaLeuArgSerAsnTyrValLeuGlyHisArgAspValGlnArgThrLeuSer 173
 |||||
 DB 216 GGAGGCCCTAAGTCCAACTATGTGCTCAAGAGACACGGGAGTGTGAGGTACACTCTCT 157
 QY 174 ProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTyrProHisTyrArgSerPro 191
 |||||
 DB 156 CAGGCAACCCACTTATACCTCATCCAGAAATGGCCACANTACCGCTCCGCC 103

RESULT 2

AW076051/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Comment

Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
 Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.llnl.gov/dbp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 252.

FEATURES

source
 1. 520
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2573424"
 /clone_1lb="NCI CGAP CML1"
 /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
 rearrangement positive, includes both chronic phase and
 myeloid blast crisis"
 /lab_host="DH10B"
 /note="Organ: whole blood; Vector: PCMV-SPORT6; Site: 1;
 Salt: Site 2: NotI; Cloned unidirectionally. Primer:
 Oligo dT. Library constructed by Life Technologies."
 BASE COUNT 107 a 139 c 160 g 108 t 6 others
 ORIGIN

Alignment Scores:

Pred. No.: 4.58e-66 Length: 520
 Score: 692.00 Matches: 129
 Percent Similarity: 89.04% Conservative: 1
 Best Local Similarity: 88.36% Mismatches: 15
 Query Match: 65.84% Indels: 1
 DB: 10 Gaps: 0

US-09-462-625-4 (1-191) x AW076051 (1-520)

QY 47 LeuProLeuArgTyrValValValSerHisThrAlaGlySerSerCysAsnThrProAla 66
 |||||
 DB 520 CTGGCTTACCGTATGTGTGTGTGATGGAACAGGCTGGCCAGTACGCTACACCTCGCCT 461
 QY 67 SerCysGlnGlnGlnAlaArgAsnValGlnHisThrMetLeuThrLeuGlyTTPCys 86
 |||||
 DB 460 TTGTGCACACACAGGCCGCGAATGTGTACACTTCCATCAATGAACACTGGGTGGTGC 401
 QY 87 AspValGlyTyrAsnPhelauIleGlyIuAspGlyLeuValTyrGlnGlyArgGlyTyr 106
 |||||
 DB 400 GACGTGGGCTTAACACTTCTGATTTGAGAAAGACGGGCTGTATACGAGGCGCTGGCTG 341
 QY 107 AsnPherHrgIuAlaHisSerGlyHisLeuThrAsnProMetSerIleGlyIleSerPhe 126
 |||||
 DB 340 AACTTCAGGGGCGCCACTCAGGTCTTATGAAACCCCATGTCATTTGCGCATTCAGCTTC 281
 QY 127 MetGlyAsnTyrMetAspArgValProThrProGlnAlaIleArgAlaIleGlnGlyLeu 146

QY	21	AspProAlaCysCysSerProLevalProArgasnclurPrpysAlaLeuAlaSerGlu	40
Db	55	-----TGCATTTTCATGTCGCCCGGACAGTGAAGAGGAGGCGCCATCCAG	102
QY	41	CysAlaGlnHisLeuSerLeuProLeuArgTyTyValValSerHisThrAlaGlySer	60
Db	103	TGCTCTGAGCCCGCTGGGCGGACAGTTCGCGTACACTTCTTTTGGAGAGGACGGTATGTC	162
QY	61	SerCysAsnThrProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyRHisMet	80
Db	163	TTCTGCACACGCCCGGACTCTCTGAAACACAGGCCCGCAATGTGCAGCATTCACACAG	222
QY	81	LysThrLeuGlyIleTPCysAspValGlyTyRAsnPhenLeuIleGlyAsnAspGlyLeuVal	100
Db	223	AATGAGCTGGGCTGTGTGCATGTAGCCTTACACTTCTTTTGGAGAGGACGGTATGTC	282
QY	101	TyrGlnGlyArgGlyIleThrAsnPhenIleGlyAlaHisSerGlyHisLeuTrpAsnPromet	120
Db	283	TATGAAAGGCGGAGCGGGAACATCAAGGGGACACACAGGCCCATCTGATCCATG	342
QY	121	SerIleGlyIleSerPheMetGlyAsnTyRMetAspArgValProThrProGlnAlaIle	140
Db	343	TCTATGGCATACCTTCATAGGGAACCTTCATGACCGGCTACCCGCAAGCGGCGCTC	402
QY	141	ArgAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyR	160
Db	403	CGGCGCCCTTAATTTCTGCAATGTGGGGTCTCGGGGCTTCTTGAGATCCACTAT	462
QY	161	ValLeuIleGlyHisArgAspValGlnArgTyRLeuSerProGlyAsnGlnLeuTyRHis	180
Db	463	GAAGTCAAAAGGACACCGGATGTGCAACACACTCTCTCCAGGTGACCAACTCTATCAG	522
QY	181	LeuIleGlnAsnTrpProHisTyRArg	189
Db	523	GTCATCCAAAGCTGGGACACATCCGA	549
RESULT 4			
LOCUS	B1149595	637 bp	mRNA
DEFINITION	602848402p1 NCI_CGAP_Lu29	Mus musculus	cDNA clone IMAGE:5011897 5'
ACCESSION	B1149595		
VERSION	B1149595.1	GI:14609596	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 637)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LRAM11062 row: c column: 02 High quality sequence stop: 632. Location/Qualifiers 1. 637 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:5011897" /clone_lib="NCI_CGAP_Lu29" /tissue_type="spontaneous tumor, metastatic to mammary." /stem_cell_origin="		


```

DB      478 GAAGTCAAGACACCGGATGTGCAACACCTCTCTCCAGGTACCAACTCTATCAG 537
OY      181 LeuileglastrProhistratg 189
DB      538 GTCATCCAAAGCTGGCAACACTACCGA 564

RESULT 6
AK008335
LOCUS   AK008335
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010107E11:peptidoglycan recognition
            protein, full insert sequence.
ACCESSION AK008335
VERSION   AK008335
KEYWORDS  HTC; CAP trapper.
SOURCE    Mus musculus (strain: C57BL/6J) adult male small intestine cDNA to
            mRRA, clone: lib:RIKEN full-length enriched mouse cDNA library
            clone:2010107E11.
ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS  Carninci, P. and Hayashizaki, Y.
TITLE    High efficiency full-length cDNA cloning
JOURNAL  Meth. Enzymol. 303, 19-44 (1999)
MEDLINE  99279253
PUBMED   10349636

REFERENCE
AUTHORS  2
TITLE    Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
JOURNAL  Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
MEDLINE  Normalization and subtraction of cap-trapper-selected cDNAs to
PUBMED   Prepare full-length cDNA libraries for rapid discovery of new genes
            Genome Res. 10 (10), 1617-1630 (2000)
            20499374
            11042159

REFERENCE
AUTHORS  3
TITLE    Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
JOURNAL  Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
MEDLINE  Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
PUBMED   Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ougaki, E., Watahiki, M.,
            Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
            RIKEN Integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
            Genome Res. 10 (11), 1757-1771 (2000)
            20530913
            11076861

REFERENCE
AUTHORS  4
TITLE    Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
JOURNAL  Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
MEDLINE  Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
PUBMED   Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
            Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
            Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
            Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
            Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M.,
            Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
            Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
            Carninci, P., de Bona, M. F., Brownstein, M. J., Bulc, C.,
            Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
            Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
            Marchionni, L., Mashima, D., Mazzarelli, J., Mombauts, P., Nordone, P.,
            Ring, B., Ringwald, C., Rodriguez, I., Sakamoto, N., Sasaki, H.,
            Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
            Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
            Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.
            and Hayashizaki, Y.
            Functional annotation of a full-length mouse cDNA collection
            Nature 409 (6821), 685-690 (2001)
            21085660
            11217851

```

```

REFERENCE
AUTHORS  5 (bases 1 to 682)
          Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
          Arkawa, T., Balderelli, R., Bono, H., Brownstein, M., Bulc, C.,
          Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Haneagaki, T.,
          Hara, A., Hayatsu, N., Hill, D., Hirokawa, K., Hirokawa, T., Hori, F.,
          Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
          Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyama, S.,
          Kurihara, R., Matsuyama, T., Miyazaki, Y., Nishi, K., Nomura, K.,
          Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
          Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
          Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
          Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
          Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamataka, I.,
          Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
          Hayashizaki, Y.
TITLE    Direct Submission
JOURNAL  Submitted (10-JUL-2000) Yoshitake Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
          URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
          Fax: 81-45-503-9216)
          Please visit our web site (http://genome.gsc.riken.go.jp/) for
          further details.
FEATURES
SOURCE    1..682
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="RANTOM:DB:2010107E11"
            /db_xref="MGD:MGI:1910652"
            /db_xref="taxon:10090"
            /clone="2010107E11"
            /sex="male"
            /tissue_type="small intestine"
            /clone_lib="RIKEN full-length enriched mouse cDNA library"
            /dex_stage="adult"
            1..682
            /gene="Pglyrp"
            38..586
            /gene="Pglyrp"
            /note="data source:MGD, source key:MGI:1345092,
            peptidoglycan recognition protein
            putative"
            /codon_start=1
            /protein_id="BAB25611.1"
            /db_xref="GI:12842460"
            /db_xref="MGD:MGI:1345092"
            /translation="MLPACALLMLDLGLATSCSFTVPSREKRALPSSCSSRLGHPVRY
            VISHRTGSCFNSPDSCEQOARNOVQHKNLGCDAVNFLLIGDGGVYEGRGVING
            DRHPTWNPISIGITFMGNFMDFPAKRALRALNLNLECGVSRGFLRSNYEVGHRDQ
            QSTLSPGDQYQYQVIOSEHYRE"
            666..671
            /gene="Pglyrp"
            /note="putative"
            682
            /gene="Pglyrp"
            /note="putative"
BASE COUNT 147 a 217 c 175 g 143 t
ORIGIN
polyA_site
polyA_signal
polyA_signal
polyA_site
682
/gene="Pglyrp"
/note="putative"
682
/gene="Pglyrp"
/note="putative"
Alignment Scores:
Pred. No.: 7 02e-66 Length: 682
Score: 692.00 Matches: 127
Percent Similarity: 78.31% Conservative: 21
Best Local Similarity: 67.20% Mismatches: 33
Query Match: 65.84% Indels: 8
DB: 11 Gaps: 1
US-09-462-625-4 (1-191) x AK008335 (1-682)
OY      1 MetleuleuAlaTTPyAlaLeuProSerLeuLeuArgLeuClYlaAlaIaGInGluThrGlu 20

```

Db	Accession	Source	Organism	Reference	Title	Journal	Comment
Db	38	ATGTTGTTGGCTGACGCTCTCTCCCTCCGAGGGCTGCGAACCTCC-----	85				
Qy	21	AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu	40				
Db	86	-----TGCAGTTTCATCTCGCCCGCCGACGTGAGTGAGGCGCCCTGCCATCCAG	133				
Qy	41	CysAlaGlnHisLeuSerLeuProLeuArgTrpValValValSerHisThrAlaIys	60				
Db	134	TGCTTACCGCCCTGGGACACCCATGTGGCTACGTGTGTATGTACACACACGCGGACG	193				
Qy	61	SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTrpHisMet	80				
Db	194	TTCTGCACACAGCCCGGACCTCGTGTACACAGCAGCGCCGCAATGTGCACATTCACACAG	253				
Qy	81	LysThrLeuGlyTrpCysAspValGlyTrpAsnPheLeuIleGlyGlnAspGlyLeuVal	100				
Db	254	AATGACGCTGGGTGGTGGATGTACCTTCACTTCATTATGGAGAGACGCTCATGTGC	313				
Qy	101	TyrGlnGlyArgGlyTrpAsnPheThrGlyAlaHisSerGlyHisLeuTrpAsnProMet	120				
Db	314	TATGAAAGCCGAGGCTGGAACTCAAGGCTGACCAACAGAGGCGCCATGTGATCCCATG	373				
Qy	121	SerIleGlyIleSerPheMetGlyAsnTrpMetAspArgValProThrProGlnAlaIle	140				
Db	374	TCTATTGGCATCACCCTTCATGTGGGAACCTTCACAGAACCGGGTACCGCAAGCGGCCCTC	433				
Qy	141	ArgAlaValGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTrp	160				
Db	434	CGTGTGCCCTTAATCTTCTGTGGAATGTGGGGGTCTCGGGGGCTTCCTGAGATCCAACTAT	493				
Qy	161	ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTrpHis	180				
Db	494	GAAAGCAAGGACACCGGGATGTGCAAGACACTCTCTCCAGGTGAGACCACTCATGAG	553				
Qy	181	LeuIleGlnAsnTrpProHisTrpArg	189				
Db	554	GTCAATCCAAAGCTGGGAACACTACCGA	580				
RESULT 7							
LOCUS	B0951956						
DEFINITION	AGENCOURT 8951306 NCI_CGAP_Co24 Mus musculus cDNA clone						
IMAGE:6474853 5', mRNA sequence.							
ACCESSION	B0951956						
VERSION	B0951956.1						
KEYWORDS	EST.						
SOURCE	house mouse.						
ORGANISM	Mus musculus.						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
TITLE	1 (bases 1 to 686)						
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.						
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)						
	Unpublished (1995)						
	Contact: Robert Strausberg, Ph.D.						
	Email: cga@bbs-remail.nih.gov						
	Tissue Procurement: The Cepko laboratory						
	cDNA Library Preparation: Life Technologies, Inc.						
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)						
	DNA Sequencing by: Agencourt Bioscience Corporation						
	Clone distribution: MGC clone distribution information can be						
	found through the I.M.A.G.E. Consortium/LMNL at:						
	http://image.llnl.gov						
	Plate: LLM14012 row: k column: 14						
	High quality sequence stop: 589.						
FEATURES							
source							
	1. 686						
	/organism="Mus musculus"						
	/strain="FVB/N"						
	/db_xref="taxon:10090"						
	/clone="IMAGE:6474853"						
	/location="NCI_CGAP_Co24"						

[illegible]

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LRAM10326 row: c column: 24
High quality sequence stop: 587.
Location/Qualifiers

FEATURES
source

1. 688
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4485071"
/clone_id="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 157 a 209 c 180 g 142 t
ORIGIN

Alignment Scores:

Pred. No.: 7.12e-66 Length: 688
Score: 692.00 Matches: 127
Percent Similarity: 78.31% Conservative: 21
Best Local Similarity: 67.20% Mismatches: 33
Query Match: 65.84% Indels: 8
DB: 12 Gaps: 1

US-09-462-625-4 (1-191) x BG244455 (1-688)

OY 1 MetLeuLeuAaTPpAlaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGlnLurHrgu 20
DB 19 ATGTGTTTGGCTGCTGCTCTCCCTCCCTGGGTCTGGCAACCTCC-----66
OY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
DB 67 -----TGCAGTTTCATCTGCTGCCCGCAGTGAAGTGAAGGCGCCCTGCCATCCGAG 114
OY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyValValValSerHisThrAlaGlySer 60
DB 115 TGCCTACGCCCGCTGGGCGACCACTGCGCTACGTGATCTCACACACACCGCGCAC 174
OY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHsMet 80
DB 175 TTCGACACAGCCGAGCTGACATCAAGGGTGAACACACAGGGCCCATCTGGAATCCATG 234
OY 81 LysThrLeuGlyTrpCysAspValGlyTyTrpAsnPhaLeuIleGlyGluAspGlyLeuVal 100
DB 235 AATGAGCTGGGCTGGTGGATGATGACCTTCAACCTTCTTATTTGAGAGGACGCTCATGTC 294
OY 101 TyrGlnGlyArgGlyTrpAsnPhaThrGlyAlaHisSerGlyHisLeuTrpAsnPromet 120
DB 295 TATGAAAGCCGAGCTGGAACATCAAGGGTGAACACACAGGGCCCATCTGGAATCCATG 354
OY 121 SerIleGlyIleSerPheMetCysAsnTyTrpMetAspArgValProThrProGlnAlaIle 140
DB 355 TCTATTGGCATCCTCTCATGGGAACTTCATGAGCCGGGTACCGCAAAAGGGGCGCTCC 414
OY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTy 160
DB 415 CGTGTGCGCCCTAAATCTTCTGGAATGTGGGGTCTCTCGGGGCTCTTGGAGATCCAACTAT 474
OY 161 ValLeuTyGlyHisArgArgValGlnArgThrLeuSerProGlyAsnGlnLeuTyHis 180
:::|||||

DB 475 GAAGTCAAAAGCACCAGGAGTGTGCAGACACTCTCTCCAGGTGACCACTATATCAG 534
OY 181 LeuIleGlnAsnTrpProHisTyrArg 189
DB 535 GTATCCAAAGCTGGGACACTACCA 561
:::|||||

RESULT 9
B1453419 703 bp mRNA linear EST 21-AUG-2001
LOCUS B1453419
DEFINITION 603170810P1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5250362.5',
mRNA sequence.
B1453419
ACCESSION B1453419
VERSION B1453419.1 GI:15244075
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Lotmar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LRAM1632 row: c column: 03
High quality sequence stop: 686.
Location/Qualifiers

FEATURES
source

1. 703
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5250362"
/clone_id="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lotmar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 163 a 214 c 179 g 147 t
ORIGIN

Alignment Scores:

Pred. No.: 7.36e-66 Length: 703
Score: 692.00 Matches: 127
Percent Similarity: 78.31% Conservative: 21
Best Local Similarity: 67.20% Mismatches: 33
Query Match: 65.84% Indels: 8
DB: 13 Gaps: 1

US-09-462-625-4 (1-191) x B1453419 (1-703)

OY 1 MetLeuLeuAaTPpAlaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGlnLurHrgu 20
DB 45 ATGTGTTTGGCTGCTGCTCTCCCTCCCTGGGTCTGGCAACCTCC-----92
OY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
DB 93 -----TGCAGTTTCATCTGCTGCCCGCAGTGAAGTGAAGGCGCCCTGCCATCCGAG 140
OY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyValValValSerHisThrAlaGlySer 60
DB 141 TGCCTACGCCCGCTGGGCGACCACTGCTGATGTCATCACACACACCGCGGAGC 200
:::|||||

```

Oy 61 SerCysasnThrProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyrHisMet 80
    |||||.....|
Db 201 TTCGCAACAGCCGACCTGCTGTAACACAGCCGCAATGTCAGCATTTACCAAG 260
Oy 81 LysThrLeuGlyTrrPysAspValGlyTyrAsnPhenLeuIleGlyLysAspGlyLeuVal 100
    |||||.....|
Db 261 AATGACCTGGCGTGGCGATGTAAGCTACACTCTCTTTTGGAGAGGAGCGGTCATGTC 320
Oy 101 TyGluGlyArgGlyTrrPasnPhenThrGlyAlaHisSerGlyHisLeuTrrPasnProMet 120
    |||||.....|
Db 321 TATGAAGCCGAGGCTGGAACATCAAGGTCAGCCACAGGCCCATCTGGAATCCCATG 380
Oy 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
    |||||.....|
Db 381 TCTATTGGCATCACTCATGAGGGAACCTTCATGAGCCGCGTACCCGCAAGCGGCGCTTC 440
Oy 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
    |||||.....|
Db 441 CGTGTGCTCCCTAATCTTCGGAATGTGGGTCTCTGGGGCTTCCTTGAATCCCACTAT 500
Oy 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
    |||||.....|
Db 501 GAAGTCAAGAGAACCGGAGATGTGCAGAAAGCACTCTCTCCAGGTGACCAACTATCAG 560
Oy 181 LeuIleGlnAsnTrrProHisTyrArg 189
    |||||.....|
Db 561 GTCATCCAAAGCTGGGAACACTACCGA 587

RESULT 10
B1155774 650 bp mRNA linear EST 05-JUL-2001
LOCUS 602904343F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5033867 5'
DEFINITION mRNA sequence.
ACCESSION B1155774
VERSION B1155774.1 GI:14615775
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 650)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1094 row: f column: 12
High quality sequence stop: 637.
Location/Qualifiers
1. 650
/organism="Mus musculus"
/strain="129,C57BL/6J,FVB/N"
/BD_xref="taxon:10080"
/clone="IMAGE:5033867"
/clone_lib="NCI_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22: 37-43 (1999)."
BASE COUNT 137 a 199 c 179 g 135 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1,08e-65 Length: 650
Score: 690.00 Matches: 127
Percent Similarity: 78.31% Conservative: 21
Best Local Similarity: 67.20% Mismatches: 33
Query Match: 65.65% Indels: 8
DB: 13 Caps: 1

US-09-462-625-4 (1-191) x B1155774 (1-650)
Oy 1 MetLeuLeuAlaTrrPalaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGlyThrGlu 20
    |||||.....|
Db 71 ATGTTGTTGGCTGCTGCTCTCTCTGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 118
Oy 21 AspProLacCysSerProIleValProArgGlnGlyTrrPylAspAlaLeuAlaSerLys 40
    |||||.....|
Db 119 -----TGCAGTTTCATGTCGCCCGCAGTGTGTGAGGCGCTCCATCCAG 166
Oy 41 CysAlaGlnHisLeuSerLeuProLeuArgTyrValAlaValSerHisThrAlaGlySer 60
    |||||.....|
Db 167 TGTCTAGCCGCTCGGGGACACCAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 226
Oy 61 SerCysasnThrProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyrHisMet 80
    |||||.....|
Db 227 TTCGCAACAGCCGACCTCTGTAACAGCAGCCGCCAATGTGCACATTTACCAAG 286
Oy 81 LysThrLeuGlyTrrPysAspValGlyTyrAsnPhenLeuIleGlyLysAspGlyLeuVal 100
    |||||.....|
Db 287 AATGACCTGGCGTGGTGGAGATGCTGTAACACTTCTTATTTGAGAGGAGCGTCATGTC 346
Oy 101 TyGluGlyArgGlyTrrPasnPhenThrGlyAlaHisSerGlyHisLeuTrrPasnProMet 120
    |||||.....|
Db 347 TATGAAGCCGAGGCTGGAACATCAAGGTCACACAGGCCCATCTGGAATCCCATG 406
Oy 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
    |||||.....|
Db 407 TCTATTGGCATCACTCATGAGGGAACCTTCATGAGCCGCTACCCGCAAGCGGCGCTTC 466
Oy 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
    |||||.....|
Db 467 CGTGTGCTCCCTAATCTTCGGAATGTGGGTCTCTGGGGCTTCCTTGAATCCCACTAT 526
Oy 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
    |||||.....|
Db 527 GAATCTCAAGAGAACCGGAGATGTGCAGAAAGCACTCTCTCCAGGTGACCAACTATCAG 586
Oy 181 LeuIleGlnAsnTrrProHisTyrArg 189
    |||||.....|
Db 587 GTCATCCAAAGCTGGGAACACTACCGA 613

RESULT 11
B1154844 610 bp mRNA linear EST 05-JUL-2001
LOCUS 602902826F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5032278 5'
DEFINITION mRNA sequence.
ACCESSION B1154844
VERSION B1154844.1 GI:14614845
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 610)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

```


Accession	Protein	Length
Qy	Tyrlguglyuaysglyttrpsnphenhrcilalyahissserglyhisleutrpasnpromt	120
Db	TATGAAGGCCGACGGTGGAACTCAATCAAGGTGTACACACACAGGCCCATCTGGAAATCCCAAG	353
Qy	Serlleglylleserphemetclvasntyrmetasparvalprothrproglinalaile	140
Db	TCATATGGCAATCACCCTTCATGGGAACCTCATGAGCCGGGTCCCGCAAGCGGGCCCCC	413
Qy	Argalaalaglnglyleuleualacysglyvalalaglnglylalaaleuargserasntyr	160
Db	CGTGTGCCCCAATACCTTCTCGAATGTGGGGGTGTCTCGGGGCTTTTGGATCCAACTAT	473
Qy	Valleuylsglyhsarqasparvalcilmrgrthleuseerproglylasnglnleutyrlis	180
Db	GAAAGTCAAAGGACACCGGAGATGTGCAAAGACACTCTCTCCAGGTGACCAACTATACG	533
Qy	Leutleuglin-Asntprprohisttyrarg	189
Db	GTCAATCCAAAGCGCTGGGAACACTACCGA	561

RESULT	13
AA228200	
LOCUS	627 bp mRNA linear EST 24-FEB-1997
DEFINITION	SMEFCA1124SK Brugia malayi microfilaria cDNA (S9W54LS-Bmf1) Brugia
ACCESSION	malayi cDNA clone SMEFCA1124 5', mRNA sequence.
VERSION	AA228200
KEYWORDS	AA228200.1 GI:1849758
SOURCE	EST.
ORGANISM	Brugia malayi1. Brugia malayi1

REFERENCE	1 (pages 1 to 627)
AUTHORS	Williams, S.A.
TITLE	Genes expressed in microfilaria of Brugia malayi
JOURNAL	Unpublished (1995)
COMMENT	Contact: Steven A. Williams

FEATURES
source
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pb1uescript SK.
Location/Qualifiers
1. .627

BASE COUNT 130 a 216 g 124 t 1 others
ORIGIN
/organism="Brugia malayi"
/strain="TNS Labs"
/db_xref="taxon:6279"
/clone="SMFCA1124"
/clone_lib="Brugia malayi microfilaria cDNA (SAM94LS-Bmmf
)"
/lab_host="X11-Blue MRF"
/note="vector: lambda unizap XR; Site.1: EcoR I; Site.2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from microfilariae of Brugia malayi
isolated from birds and converted to double stranded cDNA
using reverse transcriptase and oligo(dT) followed by
Rnase H and DnApol I. The library had 3.5 x 10⁵
independent recombinants and average insert size was 900
base pairs. The library was constructed by Lori Saunders
The library is available from Dr. S.A. Williams, email
genome@smilb.edu."

Alignment Scores:	
Pred. No.:	2.76e-64
Score:	677.00
Percent Similarity:	78.84%
Best Local Similarity:	66.67%
Query Match:	64.41%
Length:	621
Matches:	120
Conservative:	23
Mismatches:	32
Indels:	9

DB:	9	Gaps:	1
US-09-462-625-4 (1-191) x AA228200 (1-627)			

```

Oy      1 MelenleuAATPpAlaleuProSerleuNeuArgleuGlyAlaAlaGlnIuThGlu 20
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      8 ATGCTGTCGCGCTGGGCTCTCTGCGCCCTCTCTGGGCGTGGACAGCCCC----- 55

Oy      21 AspProAlaCysCysSerProIleValProArgAsnGluTrpTyrAlaLeuAlaSerGlu 40
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      56 -----TTCACCTTCGTCGTCGCCCGACGACAGTGGAGGCGCCCTCGCTGAG 103

```

Qy	Db	Qy	Db
41	104	61	164
CysAlaGlnHisIleuSerIeuProIleuAlaGlyTyrValValValSerHisIstPrAlaIleGlySer	TGCTCCAAAGCCCCCTGCACACACACACAGTCCCTACTCTGTGTATCTCACACACACAGCCAGC	SerGysasnTrpProAlaSerGysGlnGlnAlaIleArgValValGlnHisIstYrHisMet	ATCTCACACACACACAGCCCTCTCTGGAGAGAGAGAGCTGTGCATGTGGCAGCATTTACCATATG
60	153	80	223

Qy	81	LysThrLeuGlyTTPCyAspValG1TyrAspPheLeuIecTylGuaSpGlyLeuVal	1000
Db	224	AGCTGGCTGGGCTGTGTGGAGTGGCCCTACACTTCTCTATGGAGGATGGCCATGTC	2833
Qy	101	TyrGluGlyArgGlyTTPAspPheThrGlyAlaHisSerGlyHisLeuTTPAspPromet	1200
Db	284	TATGAGGCCCAAGCTGGAGTCATCAAGGATGACCCACAGGGCCACCTGTGAACCCCTTA	3433

OY	121	SerIleGGlyIeserPhemetGlyAsnTyrMetaspArgAlaProthrProgInAlaIle	140
Db	344	TCTATCGGCACACCTTCATGGGAACTACTGGATGGGTACCCCAAGCGGGCCCTC	403
OY	141	ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr	160
Db	404	CGTGTGCCCGAAGATCTTCTGGATGTGGGGTGTCTCAAGGCTTCCCGAATCCCAACTAT	463

OY	161	VallLeuTsgIyHnIsArGaSpvAlGlaATgThrLeuSeErProGlYaInglInLeuTynHis	180
		:::::	
Dd	464	GAGATCAAAAGACACCGGACTT-CATAAAAACCTTCCTCCAGCACAACGGCTTAGAT	522
		:::::	
OY	181	LeuIleGlnAsnTrpPrOnHsTyArg	189
		:::	
Dd	523	CTTAATCCGAGGCTGGGCACATAACGCC	549

RESULT	14
BG975104	
LOCUS	645 bp mRNA linear EST 12-JUN-2001
DEFINITION	602843369P1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4978057 5'
ACCESSION	BG975104
VERSION	BG975104.1 GI:14362741

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 645)
TITLE	Nih-MGC http://mgc.nci.nih.gov/ .
	National Institutes of Health. Mammalian Gene Collection (MGC)

JOURNAL COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., ritscilla.furth@ph.d
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: <http://image.llnl.gov>
Plate: LLAM10976 row: j column: 18
High quality sequence start: 3
High quality sequence stop: 654.


```
QY 88 ValGlyTyrAsnPheLeuIleGlyGluAspGlyLeuValTyrGluArgGlyTyrPasn 107.
DB 404 GTAGCCTACAACTTCTTATGTGAGAGACGTCATGTCATGTAAGCGGAGGCTGGAC 345
QY 108 PheThrGlyAlaHisSerGlyHisLeuTyrPasnPrometSerIleGlyIleSerPheMet 127
DB 344 ATCAAGGGTGACCCACACAGGGCCCATCTGGAATCCCATGCTATGTGGCATCACCTTCATG 285
QY 128 GlyAsnTyrMetAspArgValProThrProGlnAlaIleArgAlaAlaGlnGlyLeu 147
DB 284 GGAACCTTCATGTGACCGGGGACCCGCAAAAGCGGGCCCTCCGCTGCGCTTAAATCTTCTG 225
QY 148 AlaCyAGlyValAlaGlnGlyAlaLeuArgSerAsnTyrValLeuLysGlyHisArgAsp 167
DB 224 GAATGTGGGGGTGTCTCGGGCTTCTTGAGATCCAACTATGACATCAAGGACACCGGAT 165
QY 168 ValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTyrProHis 187
DB 164 GTGCAAAAGCACTCTCTCCAGGTGACCAACTCTATCAGGTCAATCCAAAGCTGGGAACAC 105
QY 188 TyrArg 189
DB 104 TACCGA 99
```

Search completed: November 12, 2002, 05:42:41
Job time : 2496 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 00:42:09 : Search time 2472 Seconds
(without alignments)
4704.029 Million cell updates/sec

Title: US-09-462-625-3

Perfect score: 718
Sequence: 1 ctgagttacggccagag.....ctccttcacataagatg 718

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estlum:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_ylt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	507.8	70.7	575	10	AM467554 he21d02.x
2	467.2	65.1	520	10	AM076051 AM076051 xa83g01.x
3	350.8	48.9	409	11	BG187105 BG187105 RST6088 A
4	316.6	44.1	682	11	AK008335 Mus muscu
5	316	44.0	703	13	BI453419 BI453419 603170810
6	315	43.9	589	12	BG174272 BG174272 602334571

7	315	43.9	610	13	BI154844
8	315	43.9	637	13	BI149595
9	315	43.9	650	13	BI155774
10	315	43.9	676	13	BI650838
11	315	43.9	686	14	BQ951956
12	315	43.9	688	12	BG244455
13	314.6	43.8	842	13	BI409815
14	314	43.7	627	9	AA228300
15	306.6	42.7	645	13	BG975104
16	306	42.6	614	10	BE199698
17	305.4	42.5	600	12	BG871384
18	304.8	42.5	555	12	BE076851
19	303	42.2	619	12	BF032505
20	303	40.8	706	12	BF163190
21	293	40.8	1045	12	BE913706
22	292.8	40.8	580	13	BI556032
23	292.2	40.7	575	9	AI507116
24	285.8	39.8	601	9	AA238752
25	277	38.6	513	14	BQ564677
26	274.2	38.2	591	9	AA238564
27	261.8	36.5	528	9	AA734893
28	259.4	36.1	547	9	AA689693
29	256.6	35.7	534	9	AA734805
30	252.8	35.2	548	9	AI585767
31	247.8	34.5	548	9	AA597240
32	240.6	33.5	384	10	BE654595
33	239	33.3	438	13	BI849729
34	236.4	32.9	518	10	BE135520
35	221	30.8	510	12	BF152060
36	215.8	30.1	384	9	AA689633
37	202.6	28.2	573	9	AA543570
38	201.4	28.1	564	9	AA499842
39	196.4	27.4	238	12	BE901379
40	180.8	25.2	451	12	BE151994
41	176	24.5	504	12	BE548605
42	175.4	24.4	472	13	BI291182
43	175.4	24.4	506	9	AA875213
44	169.2	23.6	665	13	BG969181
45	155.4	21.6	384	9	AA015317

ALIGNMENTS

RESULT 1
AM467554/c 575 bp mRNA linear EST 24-FEB-2000
LOCUS he21d02.x1 NCI-CGAP CML1 Homo sapiens cDNA clone IMAGE:2919651 3'
DEFINITION similar to TR:O75594 O75594 PEPTIDOGLYCAN RECOGNITION PROTEIN
PRECEDSOR. ; mRNA sequence.

ACCESSION AM467554
VERSION AM467554.1 GI:7037660
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 575)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Seq primer: -40UP from GIBCO

High quality sequence stop: 359.

FEATURES

Location/Qualifiers

1..575

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="NCI_CGAP-CMU1"

/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL rearrangement positive, includes both chronic phase and myeloid blast crisis"

/lab_host="DH10B"

/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt; Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

BASE COUNT 110 a 147 c 178 g 135 t 5 others

ORIGIN

Query Match 70.7%; Score 507.8; DB 10; Length 575;

Best Local Similarity 95.8%; Pred. No. 3.9e-108;

Matches 529; Conservative 0; Mismatches 22; Indels 1; Gaps 1;

Location/Qualifiers

1..520

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="NCI_CGAP-CMU1"

/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL rearrangement positive, includes both chronic phase and myeloid blast crisis"

/lab_host="DH10B"

/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt; Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

BASE COUNT 107 a 139 c 160 g 108 t 6 others

ORIGIN

Query Match 65.1%; Score 467.2; DB 10; Length 520;

Best Local Similarity 95.1%; Pred. No. 1.1e-98;

Matches 489; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

Location/Qualifiers

1..520

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="NCI_CGAP-CMU1"

/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL rearrangement positive, includes both chronic phase and myeloid blast crisis"

/lab_host="DH10B"

/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt; Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

BASE COUNT 107 a 139 c 160 g 108 t 6 others

ORIGIN

Query Match 65.1%; Score 467.2; DB 10; Length 520;

Best Local Similarity 95.1%; Pred. No. 1.1e-98;

Matches 489; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

Location/Qualifiers

1..520

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="NCI_CGAP-CMU1"

/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL rearrangement positive, includes both chronic phase and myeloid blast crisis"

/lab_host="DH10B"

/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt; Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

RESULT 2
 AM076051/c 520 bp mRNA linear EST 13-OCT-1999
 LOCUS AM076051 NC_000001.1 Homo sapiens cDNA clone IMAGE:2573424 3'
 DEFINITION x883901.x1 NCI_CGAP CMU1 Homo sapiens cDNA clone IMAGE:2573424 3'
 PRECURSOR: // mRNA sequence.
 ACCESSION AM076051
 VERSION AM076051.1 GI:6031049
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 520)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

AUTHORS

Contact: Robert Strausberg, Ph.D.
 Email: cgapdb-remail.nih.gov
 Tissue Procurement: Elisabeth Paletta, Jonathan D. Licht, M.D.,
 Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
 Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrip/image/image.html
 Seq primer: -400p from Glibco

TITLE

High quality sequence stop: 252.
 Location/Qualifiers

JOURNAL

1..520
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="NCI_CGAP-CMU1"
 /tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL rearrangement positive, includes both chronic phase and myeloid blast crisis"
 /lab_host="DH10B"
 /note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt; Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

COMMENT

BASE COUNT 107 a 139 c 160 g 108 t 6 others

ORIGIN

Query Match 65.1%; Score 467.2; DB 10; Length 520;
 Best Local Similarity 95.1%; Pred. No. 1.1e-98;
 Matches 489; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

FEATURES

Location/Qualifiers

1..520

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="NCI_CGAP-CMU1"

/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL rearrangement positive, includes both chronic phase and myeloid blast crisis"

/lab_host="DH10B"

/note="Organ: whole blood; Vector: pCMV-SPORT6; Site: 1; Salt; Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

BASE COUNT 107 a 139 c 160 g 108 t 6 others

ORIGIN

Query Match 65.1%; Score 467.2; DB 10; Length 520;
 Best Local Similarity 95.1%; Pred. No. 1.1e-98;
 Matches 489; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

ORIGIN

RESULT 2
 AM076051/c 520 bp mRNA linear EST 13-OCT-1999
 LOCUS AM076051 NC_000001.1 Homo sapiens cDNA clone IMAGE:2573424 3'
 DEFINITION x883901.x1 NCI_CGAP CMU1 Homo sapiens cDNA clone IMAGE:2573424 3'
 PRECURSOR: // mRNA sequence.
 ACCESSION AM076051
 VERSION AM076051.1 GI:6031049
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Db	40	AAAAACCCACTGCTCCTTCATATAAGATG	7
RESULT_3			
LOCUS	BGI87105	409 bp	mRNA linear EST 21-APR-2001
DEFINITION	R56088 Athersys RAGE Library Homo sapiens CDNA, mRNA sequence.		
ACCESSION	BGI87105		
VERSION	BGI87105.1	GI:13708792	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 409)		
AUTHORS	Harrington,J.J., Sherif,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cuthren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.		
TITLE	Creation of genome-wide protein expression libraries using random activation of gene expression		
JOURNAL MEDLINE	Nat. Biotechnol. 19 (5), 440-445 (2001)		
COMMENT	21227151 Contact: Scott J. Cain. Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scjain@atersys.com High quality sequence stop: 409. Location/Qualifiers 1..409 .organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Athersys RAGE library" /cell_line="HT1080" /note-"See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."		
BASE COUNT	93 a 139 c 95 g 81 t		
ORIGIN			
Query Match	48.9%; Score 350.8; DB 12; Length 409;		
Best Local Similarity	99.2%; Pred. No. 1.3e-71;		
Matches	363; Conservative 0; Mismatches 2; Indels 1; Gaps 1;		
OY	354	AAGAGGGCGTGTATAGAGGGCGTGCGTAACCTCACGGGTGCCACTGAGTGCAC-	412
Db	25	AAAAGGGCGTGTATAGAGGGCGTGCGTAACCTCACGGGTGCCACTGAGTGCAC	84
OY	413	TTATGGAACCCCATGTCATTGGCATTCACGCTCAATGAGGCACTAAATGATGGGTGCC	472
Db	85	TTATGGAACCCCATGTCATTGGCATTCACGCTCAATGAGGCACTAAATGATGGGTGCC	144
OY	473	ACACCCCAAGCCATCGGGACAGCCAGGAGTCTACGCGTGGGTGGTGCACGAGACC	532
Db	145	ACACCCCAAGCCATCGGGACAGCCAGGAGTCTACGCGTGGGTGGTGCACGAGACC	204
OY	533	CTGAGTCCAATATGTGCTCAAAGACACCGGATGTGACAGCTACACTCTCTCCAGGC	592
Db	205	CTGAGTCCAATATGTGCTCAAAGACACCGGATGTGACAGCTACACTCTCTCCAGGC	264
OY	593	AACCAAGCTTACACTATATCAAGAATTGGCACTACCGTCCCCCTGAGGCCCTGCT	652
Db	265	AACCAAGCTTACACTATATCAAGAATTGGCACTACCGTCCCCCTGAGGCCCTGCT	324
OY	653	GATCGCAACCCATTCGCCCCTCCCATGGCCAAACCCCAACCGCTGCTCTCCCAATA	712
Db	325	GATCGCAACCCATTCGCCCCTCCCATGGCCAAACCCCAACCGCTGCTCTCTCCCAATA	384

QY	713	AAGATG	718
Db	385	AAGATG	390
RESULT 4			
AK008335			
LOCUS	682 bp	RNA	linear
DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length		
ACCESSION	AK008335		
VERSION	AK008335.1	GI:12842459	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain: C57BL/6J) adult male small intestine cDNA to mRNA, clone: lib. RIKEN full-length enriched mouse cDNA library		
ORGANISM	clone: 2010107E11.		
REFERENCE	Mus musculus		
ATHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1		
ATHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
ATHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
ATHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, K., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
ATHORS	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, Y., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Banno, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kiehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staab, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Botfield, D., Boujard, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, P., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Welt, C., Wittaker, C., Wilmink, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409 (6821), 685-690 (2001)		
MEDLINE	21085660		
PUBMED	11217851		
REFERENCE	5 (bases 1 to 682)		

AUTHORS
Adechi, J., Aizawa, K., Akahira, S., Akamura, T., Aono, H., Arai, A.,
Arakawa, T., Balderelli, R., Bono, H., Brownstein, M., Butt, C.,
Carcinici, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hall, D., Hiramoto, K., Hiroka, T., Horii, F.,
Hume, D., Imocanti, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koye, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishii, K., Nomura, K.,
Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schirrm, L., Shibata, K., Shibata, Y., Shihagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, I.,
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, F.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

TITLE
Direct Submission

JOURNAL
Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suheiho-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,
URI: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

FEATURES	Location/Qualifiers
source	1. .682

```

/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="PANTOM_DB:2010107E11"
/db_xref="MGD:MGI:1910652"
/db_xref="taxon:10090"
/clone="2010107E11"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..682
/gene="Pglyrp"
38..586
CDS

```

```

/gene="Pglyrp"
/note="data source:MCD, source key:MGI:1345092,
evidence:ISS
peptidoglycan recognition protein
putative"
/codon_start=1
/protein_id="BAB25611.1"
/db_xref="GI:12842460"
/db_xref="MGI:MGI:1345092"
/translition="MLFACALLALILGLNTSCFYIPRSEMRALPESGSRIGHPIRYKY
VISHAGSECPNPSDEOQARVQVHVNENEGMDVAVNELLGEGHYHTEGGKMKY
DHTGQIINPMSTGCTPMGKNEMDRVPAAKRALRAALMLLECGVSRGFLBSNTEYKGRHD
OSTLSFQDQLVIOVISWEHYRE"
666..671
polyA_signal

```

```

/gene="Pglyrp"
/note="putative"
682
polyA_site

```

ORIGIN	BASE COUNT				
147	a	217	c	175	g
					143 t

Query Match	44.18;	Score 316.6;	DB 11;	Length 682;
Best Local Similarity	77.28;	Pred. No. 1.3e-63;		
Matches 385;	Conservative 0;	Mismatches 114;	Indels 0;	Gaps 0

Accession	Sequence	Position
Oy	CGTGTGAGGCCCATATATGCCCCGGAACAGTGTGAAGGCCCTGACATCAAGAGGCCCC	194
Oy	135 CCGTGTGAGGCCCATATATGCCCCGGAACAGTGTGAAGGCCCTGACATCAAGAGGCCCC	135
Db	81 CCTCTGAGAGTTTACTGCTGCCCCGCACTAGATGAGAGGCCCTTCACATCCGAGTCTCTA	140
Oy	135 AGCACTAGACCTGCGCCCTTACGCTATATGTGTGTGTATGTCACATCGCGGGCAGACACTGCA	25
Oy	135 AGCACTAGACCTGCGCCCTTACGCTATATGTGTGTGTATGTCACATCGCGGGCAGACACTGCA	135
Db	141 GCGCGCTTGGGGACCCCACTTGCTACAGGTGGATCTCACACACACCGGGCAGCTTTCGCA	200
Db	141 GCGCGCTTGGGGACCCCACTTGCTACAGGTGGATCTCACACACACCGGGCAGCTTTCGCA	141

OY	255	ACACCCCGCCCTCGCGCCAGGACGAGGCGCGGAATGTGCAGACTACCAACATGAAGACAC	314
Db	201	ACAGCCCGGACTCCTGTGTAAACAGCAGGCCCGCAATGTGCAGCATTTACCAACAAGATTGAC	260
OY	315	TGGGCTGGTGGACGTGGGCTACAACTCTCTGATTGGAGAAAGACGGGCTCGTATACGAG	374
Db	261	TGGGCTGGTGGGATGTAGCTCTACCAACTCTCTTATTGGAGAGGACGGCTATGTCTATGAAG	320
OY	375	GCCGTGGCTGGAACTTTCACGGGTGCCACTCAGTCACTTATGAAACCCCATGTCCATTG	434
Db	321	GCCGAGGCTGGAACTTCAAAGGTTGCCACACAGGCCCATCTGGAAATCCATGTCTATTGG	380
OY	435	GCATAGCTTATGCGGCAACTACATGATGGGTGGGTCGCCACCCCAAGGCCATCGGGGAG	494
Db	381	GCATACCTTCACTGCGGGAACCTTACATGACCGGGTACCCGCACAAAGCGGCCCTCCGCTGG	440
OY	495	CCACAGGCTACTGCGCTCGCGGTGTGGGCTCAGGAGACCCCTGAGGTCCCACTATGTCTCA	554
Db	441	CCCTAAATCTTCTGGAATGTGGGGGTGTCTCGGGGCTTCTGTGATTCCAACTATGAAGTCA	500
OY	555	AAGGACACCGGGATGTGCAGCGTACACTCTCTCCACGCAACACGCTCTTACCACTTCATCC	614
Db	501	AAGGACACCGGGATGTGCAAAAGCACTCTCTCCAGGTGACCAACCTTATCAGGTATCC	560
OY	615	AGAAATGGCCACACATACCG	633
Db	561	AAAGCTGGGAACACTACCG	579

RESULT 5					
BI453419					
LOCUS					
DEFINITION	703 bp	mRNA	linear	EST 21-AUG-2001	
ACCESSION	603170810P1	NCI_CGAP_Mam5	Mus musculus	CDNA clone	IMAGE:5250362 5'
VERSION	BI453419				
KEYWORDS	BI453419.1	GI:15244075			
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Mammalia; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	1 (bases 1 to 703)	NH-MGC http://mgc.nhl.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)	Contact: Robert Strausberg, Ph.D.

Email: cgapbs-email.nih.gov
 Tissue Procurement: Lothar.Hennighausen.Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MCC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L1AM1632 row: c column: 03
 High quality sequence stop: 686.

FEATURES	Location/Qualifiers
source	1. .703

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5250362"
/clone_1db="NCI CGAP Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"

```

```

BASE COUNT      163 a      214 c      179 g      147 t
ORIGIN
/Note="Organ: mammary; Vector: PCMV-SPOR6; Site:1: Salt
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators:
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

```

```

FEATURES
  source
    . Location/Qualifiers
      1. .589
        /organism="Mus musculus"
        /strain="FVB/N"

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia, Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 610)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

555 AAGGACACCCGGGATGTCAGGCTACACTCTCTCCAGGCAACCACTCTACCACTCAATCC 614

135 CCCTGTCAGCCCATAGTGTGCCCCGGAACGAGTGGAAAGCCCTGGCATCAGAGTGGCCC 194

Db 65 CCTCTGACATTCATGTCGCCCCGACATGATGAGAGGCCCTGCATCCGATGCTCTA 124
 QY 195 AGACCTGAGCGCTGCGCTTACGCTATGTGGTGATGGCAACAGCGGGGACGACTGCA 254
 Db 125 GCCGCTGGGGGACACCGCTGCTGATCTGATCCACACAGCGGGGCTTGCA 184
 QY 255 ACACCCCGGCTGCTGTCAGCAGAGCGCCGATGTGCACATACACATGAAGAC 314
 Db 185 ACAGCCCGGACTCCTGTGAACAGAGCGCCGATGTGCACATACACATGAAGAC 244
 QY 315 TGGGCTGTGTCGACAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAG 374
 Db 245 TGGGCTGTGTCGACAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAG 304
 QY 375 GCGGTGGCTGGAAGTGGAGGGTGGCCACATGATCTATGTAACCCCATGTGCAATG 434
 Db 305 GCCGAGGCTGGAACATGAGGGTGGACACAGGGCCCATCTGGAATCCCATGTCTATG 364
 QY 435 GCATCAGCTTCATGAGGACATGATGATGGTGGCCACACCCAGCCATCCGGGACAG 494
 Db 365 GCATCAGCTTCATGAGGACATGATGATGGTGGCCACACCCAGCCATCCGGGACAG 424
 QY 495 CCCAGGCTTACTGCGCTGCGGCTGTGCTCAGGAGCGCTGAGGTCCACTATGTCTCA 554
 Db 425 CCTTAATCTTGTGAATGTGGGTGTCTCGGGCTTCTTGAGATCCAACTATGAACTCA 484
 QY 555 AAGGACACCGGATGTGACAGCTCTCTCCAGGACACGACCTTACCACTCATCCATCC 614
 Db 485 AAGGACACCGGATGTGACAGCTCTCTCTCCAGGACACGACCTTACCACTCATCCATCC 544
 QY 615 AGAATTGGCCACTACCG 633
 Db 545 AAAGCTGGGAACACTACCG 563

RESULT 11
 BQ951956
 LOCUS BQ951956 686 bp mRNA linear EST 21-AUG-2002
 DEFINITION AGENCOURT. BQ951956 NCI_CGAP_Co24 Mus musculus cDNA clone
 IMAGE:6474853 5', mRNA sequence.
 BQ951956
 ACCESSION BQ951956.1 GI:22367434
 VERSION BQ951956.1 GI:22367434
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 686)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bcr-remail.nih.gov
 Tissue Procurement: The Cepho Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M14012 row: k column: 14
 High quality sequence stop: 589.
 Location/Qualifiers
 1..686
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image:6474853
 /clone_lib="NCI_CGAP_Co24"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
 Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 Kb. Constructed by Life

BASE COUNT 157 a 207 c 172 g 141 t 9 others
 ORIGIN
 Query Match. 43.9% Score 315; DB 14; Length 686;
 Best Local Similarity 77.0%; Pred. No. 3e-63;
 Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 135 CCTGTGACAGCCCATAGTGGCCCGGACAGTGGAGAGCCCTGTGATCAGATGCGGCC 194
 Db 63 CCTGTGACAGTGGTATGATGCGCCCGGACAGTGGAGAGCCCTGTGATCAGATGCGGCC 122
 QY 195 AGACCTGAGGCTCCCTTACGCTATGTGTGTATGTCACACAGCGGGGACGACTGCA 254
 Db 123 GCCGCTGGGGGACCCAGTTCGATGCTGATCTACACACAGCGGGGACGCTTCTGCA 182
 QY 255 ACACCCCGGCTGCTGTCAGCAGAGCGCCGATGTGCACATACACATGAAGACAC 314
 Db 183 ACACCCCGGACTCCTGTGAACAGAGCGCCGATGTGCACATTAACCAAGATGAGC 242
 QY 315 TGGGCTGTGTCGACAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAG 374
 Db 243 TGGGCTGTGTCGACAGTGGGCTTACACTCTGATTTGAGAGAGCGGCTGTATACAG 302
 QY 375 GCGGTGGCTGGAAGTGGAGGGTGGCCACATGATCTATGTAACCCCATGTGCAATG 434
 Db 303 GCCGAGGCTGGAACATGAGGGTGGACACAGGGCCCATCTGGAATCCCATGTCTATG 362
 QY 435 GCATCAGCTTCATGAGGACATGATGATGGTGGCCACACCCAGCCATCCGGGACAG 494
 Db 363 GCATCAGCTTCATGAGGACATGATGATGGTGGCCACACCCAGCCATCCGGGACAG 422
 QY 495 CCCAGGCTTACTGCGCTGCGGCTGTGCTCAGGAGCGCTGAGGTCCACTATGTCTCA 554
 Db 423 CCTTAATCTTGTGAATGTGGGTGTCTCGGGCTTCTTGAGATCCAACTATGAACTCA 482
 QY 555 AAGGACACCGGATGTGACAGCTCTCTCCAGGACACGACCTTACCACTCATCCATCC 614
 Db 483 AAGGACACCGGATGTGACAGCTCTCTCTCCAGGACACGACCTTACCACTCATCCATCC 542
 QY 615 AGAATTGGCCACTACCG 633
 Db 543 AAAGCTGGGAACACTACCG 561

RESULT 12
 BG244455
 LOCUS BG244455 688 bp mRNA linear EST 13-FEB-2001
 DEFINITION 602356590F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4405071 5',
 mRNA sequence.
 BG244455
 ACCESSION BG244455.1 GI:12754270
 VERSION BG244455.1 GI:12754270
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 688)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bcr-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M10326 row: c column: 24
 High quality sequence stop: 587.
 Location/Qualifiers

FEATURES

source

1. 688
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4485071"
/clone_1ib="NCI-CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SpOxT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 157 a 209 c 180 g 142 t

ORIGIN

Query Match 43.9%; Score 315; DB 12; Length 688;
Best Local Similarity 77.0%; Pred. No. 3e-63;
Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 135 CCTGCTGACAGCCCATATGTCGCCCGAGAGAGTGAAGGCCCTGGCATCAGTGGCC 194
DB 62 CCTCTGCAAGTTTATCAGTCGCCCCGACATGAGAGAGGCCCTGCGCATCGAGTGTCTA 121
QY 195 AGCAGCTGAGCTGCCCTTACGCTATGTGTGTATCGACAGCGGCGAGCAGCTGCA 254
DB 122 GCCGCTGGGGACACCGAGTTGCTGATCTCACACACAGCGCGAGCTTCTGCA 181
QY 255 ACACCCCGGCTGTCGTCGACAGAGCCCGGAAATGTGACGACTACCATGAGACAC 314
DB 182 ACAGCCCGGAGCTCTGTGAACACAGAGCCCGCAATGTGACGACTTACACACAGAAAGAGC 241
QY 315 TGGGCTGTGTCGACGAGTGGCTACAACTCTGATTTGAGAGAGAGGCTGATAGAGG 374
DB 242 TGGGCTGTGTCGACGAGTGGCTACAACTCTGATTTGAGAGAGAGGCTGATAGAGG 301
QY 375 GCGGTGCTGGAATCTCAGCGGTCGCCACTAGTCACTTATGGAACCCCATGCTCAT 424
DB 302 GCGAGGCTGGAACATCAAGAGGTGACACACAGGGGCCCATCTGATTCATGTCAT 361
QY 435 GCATGAGCTTCAATGGGCACTACATGATGAGTGGCTGCCACACCCGAGGCTCGGGCAG 494
DB 362 GCATGAGCTTCAATGGGCACTACATGAGTGGCTGCCACACCCGAGGCTCGGGCAG 421
QY 495 CCCAGGCTTACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 554
DB 422 CCCAAATCTTCTGGAATGTGGGGTGTCTCGGGGCTTCTGGAATCTCACTAAGATCA 481
QY 555 AAGGACACCGGATGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 614
DB 482 AAGGACACCGGATGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
QY 615 AGAATTGGCCACACTACCG 633
DB 542 AAAGCTGGGAACACTACCG 560

RESULT 13
BI409815 842 bp mRNA linear EST 14-AUG-2001
LOCUS 602961906F1 NCI-CGAP_Lu33 Mus musculus cDNA IMAGE:5117688 5',
DEFINITION mRNA sequence.
ACCESSION BI409815
VERSION BI409815.1 GI:15170738
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 842)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM1286 row: k column: 01
High quality sequence stop: 568.
Location/Qualifiers

1. 842
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5117688"
/clone_1ib="NCI-CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker. Site_1: NotI; Site_2: EcoRI. 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dt) primer [5' TTTTACCAATCTGAGAGTGGAGCGGCGCTGTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 201 a 259 c 226 g 154 t 2 others

ORIGIN

Query Match 43.8%; Score 314.6; DB 13; Length 842;
Best Local Similarity 76.8%; Pred. No. 3.8e-63;
Matches 383; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 135 CCTGCTGACAGCCCATATGTCGCCCGAGAGTGAAGGCCCTGGCATCAGTGGCC 194
DB 34 CCTCTGCAAGTTTATCAGTCGCCCCGACATGAGAGAGGCCCTGCGCATCGAGTGTCTA 93
QY 195 AGCAGCTGAGCTGCCCTTACGCTATGTGTGTATCGACAGCGGCGAGCAGCTGCA 254
DB 94 GCCGCTGGGGACACCGAGTTGCTGATCTCACACACAGCGCGAGCTTCTGCA 153
QY 255 ACACCCCGGCTGTCGTCGACAGAGCCCGGAAATGTGACGACTACCATGAGACAC 314
DB 154 ACAGCCCGGAGCTCTGTGAACACAGCGCCGCAATGTGACGACTTACACACAGATGAGC 213
QY 315 TGGGCTGTGTCGACGAGTGGCTACAACTCTGATTTGAGAGAGGCTGATAGAGG 374
DB 214 TGGGCTGTGTCGACGAGTGGCTACAACTCTGATTTGAGAGAGGCTGATAGAGG 273
QY 375 GCGGTGCTGGAATCTCAGCGGTGCCACTCAGTCACTTATGGAACCCCATGCTCAT 424
DB 274 GCGAGGCTGGAACATCAAGAGGTGACACACAGGCCCATCTGGAATCTTATAGG 333
QY 435 GCATGAGCTTCAATGGGCACTACATGATGCGTCCACACCCGAGGCTCGGGCAG 494
DB 334 GCATGAGCTTCAATGGGCACTACATGATGCGTCCACACCCGAGGCTCGGGCAG 393
QY 495 CCCAGGCTTACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 554
DB 394 CCTTAATCTTCTGGAATGTGGGGTGTCTCGGGGCTTCTGGAATCTCACTAAGATCA 453
QY 555 AAGGACACCGGATGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 614
DB 454 AAGGACACCGGATGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
QY 615 AGAATTGGCCACACTACCG 633
DB 514 AAAGCTGGGAACACTACCG 532

FEATURES	Location/Qualifiers
source	1. .627

```

/lab.host="XLI-Blue MRF"
/note="Vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from microfilariae of Brugia malayi
isolated from jirds and converted to double stranded cDNA
using reverse transcriptase and oligo(dT) followed by
RNase H and DNasep I. The library had 3.5 x 105
independent recombinants and average insert size was 900
base pairs. The library was constructed by Iori Saunders.
The library is available from Dr. S.A. Williams, email
genome@smith.edu."

```

Query Match	43.7%;	Score 314;	DB 9;	Length 627;
Best Local Similarity	77.7%;	Pred. No. 5e-63;		
Matches 391;	Conservative	0;	Mismatches 111;	Indels 1;
			Gaps	1;

Qy	132	CGGGCTGCTGAGGCCCATAGTGGCCCCGGAAACGAGTGGAAAGGCGCTGGATCAAGATGCGC	191
Db	48	CAGCCCGCTGGAGTTGTTGTGTGCCCCCGACACGAGTGGAGAGGCCCTGCCGTCTAAGTCT	107
Qy	192	CCCAAGCACTGAGCCTGCGCTTACGCTATGTGTGTATGCGACAGGGGGGAGCAGCT	251
Db	108	CCAAAGCCCTTGACACAAACCACTCCGCTACGCTGAGTCACTACACAGCAGGCGAGCAATCT	167
Qy	252	GCAACACACCCCGCGCTCGAGGCCAGAGAGGGCGGGAATGTGCACACACTACACATGAGA	311
Db	168	GCACCAAGCCAGCCTCTCTGCGAGACGAGGCTTCCGCAATGTGCACACATTAACCAATGAGCT	227
Qy	312	CACGTGGGCTGTGTGCAGCTGGGCTACAACACTCTCTGATTGGAGAAAGACGGGCTCCTATACG	371
Db	228	CGCTGGGCTGTGTGCAGCTGGGCTACAACACTCTCTCATTTGGAGAGAGATGGCCATGTATG	287
Qy	372	AGGGCGCGTGGGTGAACCTTCACAGGGGTGCCCACTAGGTCACTTTTGGAAACCCATGTCCA	431
Db	288	AGGGCCNAGGCTGTGAGCACTCAAGGTGACACACAGAGGCCACCTTGGAAACCCCTTAATCTA	347

RESULT	15
LOCUS	BG975104
DEFINITION	BG975104 645 bp mRNA linear EST 12-JUN-2001
ACCESSION	U0284336F1 NCI-GAP Mam4 Mus musculus cDNA clone IMAGE:4979057 5'
VERSION	BG975104
KEYWORDS	BG975104.1 GI:14362741
SOURCE	EST.
ORGANISM	house mouse. Mus musculus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 645)		National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	Contact: Robert Strausberg, Ph.D.

tissue Procurement: Lotmar Hennighausen Ph.D., Priscilla Furt-
 Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: HAM10976 row: j column: 18
 High quality sequence start: 3
 High quality sequence stop: 634.

FEATURES	Location/Qualifiers
source	1. .645

```

/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4979057"
/clone_l1b="NCI_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."
```

	Query Match	42.7%	Score 306.6	DB 13	Length 645
	Best Local Similarity	77.3%	Pred. No. 2,76-61		
	Matches 372	Conservative 0	Mismatches 109	Indels 0	Gaps 0
QY	135 CCTGCTGACGCCCATAGTGCCTCCCGGAACGATGAGAGGCGCTTGGCATCAGAGTCCGCC	194			
Db	109 CTCTCTGACAGTTTATCTGCGCCGACATGATGATGAGAGGCGCTTGGCATCAGAGTCTCTTA	168			

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 00:03:09 ; Search time 2720 seconds

(without alignments)
7682.285 Million cell updates/sec

Title: US-09-462-625-3

Perfect score: 718

Sequence: 1 ctgagtactgagccagag.....ctcctctccaataagatg 718

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hgt_hum: *
31: em_hgt_inv: *
32: em_hgt_other: *
33: em_hgt_mus: *
34: em_hgt_pln: *
35: em_hgt_rod: *
36: em_hgt_mam: *
37: em_hgt_vit: *
38: em_sy: *
39: em_hgt_hum: *
40: em_hgt_mus: *
41: em_hgt_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	673	93.7	724	9 AF242517	AF242517 Homo sapi
2	653.2	91.0	690	4 AF076483	AF076483 Homo sapi
3	393.6	54.8	700	4 CDR131676	AJ131676 Camelus d
4	349.4	48.7	688	4 AY083309	AY083309 Bos tauru
5	315.6	44.1	549	6 AR124884	AR124884 Sequence
6	315.6	44.1	669	10 AF193843	AF193843 Mus muscu
7	316.6	44.1	680	10 AF076482	AF076482 Mus muscu
8	315	43.9	713	10 BC005582	BC005582 Mus muscu
9	302.6	42.1	678	10 MMRNATMS1	X86374 M. musculus
10	299	41.6	166500	9 AC007785	AC007785 Homo sapi
11	295.6	41.2	630	10 AF154114	AF154114 Rattus no
12	185.4	25.8	259	4 SSC310355	AJ10355 Sus scrof
13	161.4	22.5	5358	4 CDR409286	AJ409286 Camelus d
14	147.4	20.5	166867	3 AC007085	AC007085 Drosophi
15	147.4	20.5	169509	3 AC007303	AC007303 Drosophi
16	147.4	20.5	172838	2 AC020451	AC020451 Drosophi
17	147.4	20.5	259336	3 AE003836	AE003836 Drosophi
18	145	20.2	651	3 AY060645	AY060645 Drosophi
19	145	20.2	652	3 AF207542	AF207542 Drosophi
20	136.4	19.0	717	3 AF207541	AF207541 Drosophi
21	136.4	19.0	1162	3 AY075293	AY075293 Drosophi
22	132.2	18.4	1107	6 AX119920	AX119920 Sequence
23	132.2	18.4	1125	6 AX119917	AX119917 Sequence
24	127	17.7	4205	10 MMTAG7	Y12088 Mus muscu
25	122	17.0	182897	2 AC110846	AC110846 Rattus no
26	117.6	16.4	1110	6 AX119918	AX119918 Sequence
27	117.6	16.4	1128	6 AX119915	AX119915 Sequence
28	117.6	16.4	1194	9 AY035377	AY035377 Homo sapi
29	117.4	16.4	1858	9 AF242518	AF242518 Homo sapi
30	115.8	16.1	1173	9 AY035376	AY035376 Homo sapi
31	97.8	13.6	521	6 E16478	E16478 Bombyx mori
32	97.8	13.6	590	6 E16479	E16479 Bombyx mori
33	97.8	13.6	753	6 E16480	E16480 Bombyx mori
34	96.2	13.4	735	6 AB016249	AB016249 Bombyx mo
35	94.6	13.2	182897	2 AC110846	AC110846 Rattus no
36	94	13.1	1794	9 AF384856	AF384856 Homo sapi
37	93.2	13.0	718	3 AF441723	AF441723 Bombyx mo
38	92.4	12.9	1374	3 AY058258	AY058258 Drosophi
39	92.4	12.9	1375	3 AF313391	AF313391 Drosophi
40	92.4	12.9	2751	9 AK055882	AK055882 Homo sapi
41	90.2	12.6	1369	3 AB084190	AB084190 Drosophi
42	89	12.4	670	3 AF076481	AF076481 Trichoplu
43	88.8	12.4	622	3 AB017520	AB017520 Bombyx mo
44	87	12.1	55863	2 AC018100	AC018100 Drosophi
45	87	12.1	157225	3 AC010042	AC010042 Drosophi

ALIGNMENTS

RESULT 1
AF242517
LOCUS AF242517 724 bp mRNA linear PRI 15-AUG-2000
DEFINITION Homo sapiens hypothetical protein SB168 mRNA, complete cds.
ACCESSION AF242517
VERSION AF242517.1 GI:9802032
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 724)
AUTHORS Wan, T., Zhang, W. and Cao, X.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Department of Immunology, Second Military

|||||
 Db 301 AGACACTGGGCTGTCGACGCTGCTACACTCTCTGATGTGAGAGACGGGCTGTAT 360
 369 AGGAGGGCCGTGGTGGAACTTACGGGNGCCCACTCAGTACTATATGAAACCCATAT 428
 Db 361 ACGAGGGCCGTGGTGGAACTTACGGGNGCCCACTCAGTACTATATGAAACCCATAT 420
 QY 429 CCATTGGCATGCTTCAATGGGCACTACATGATCGGGTGGCCACACCCAGGCCATCC 488
 Db 421 CCATTGGCATGCTTCAATGGGCACTACATGATCGGGTGGCCACACCCAGGCCATCC 480
 QY 489 GGGCAGCCAGGGTCTACTGCGCTCGGTGGTGTGAGGAGCCCTGAGGTCCTAATG 548
 Db 481 GGGCAGCCAGGGTCTACTGCGCTCGGTGGTGTGAGGAGCCCTGAGGTCCTAATG 540
 QY 549 TGGCTAAAGAGACGGGGTGTGAGCGGTACACTCTCTCAGGCAACAGCTCTACAC 608
 Db 541 TGGCTAAAGAGACGGGGTGTGAGCGGTACACTCTCTCAGGCAACAGCTCTACAC 600
 QY 609 TCATCAGAAATTTGGCCACACTACCGCTCCCTGAGAGCCCTGTGATCGCACCCATTC 668
 Db 601 TCATCAGAAATTTGGCCACACTACCGCTCCCTGAGAGCCCTGTGATCGCACCCATTC 660
 QY 669 CTCCCTCCCATGGCCAAAACCCCATGT 698
 Db 661 CTCCCTCCCATGGCCAAAACCCCATGT 690

RESULT 3
 CDR131676 700 bp mRNA linear NAM 21-DEC-2000
 LOCUS Camelus dromedarius mRNA for peptidoglycan recognition protein.
 DEFINITION AJ131676
 VERSION AJ131676.1 GI:11990123
 KEYWORDS peptidoglycan recognition protein.
 SOURCE Arabian camel.
 ORGANISM Camelus dromedarius
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 REFERENCE 1 (bases 1 to 700)
 AUTHORS Kappeler S.R., Farah Z. and Puhan Z.
 TITLE Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan
 Recognition Protein
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 700)
 AUTHORS Kappeler S.R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1998) Kappeler S.R., Institute of Food Science,
 Swiss Federal Institute of Technology, LFO F26, 8092 Zurich,
 SWITZERLAND

FEATURES
 source location/Qualifiers
 1..700
 /organism="Camelus dromedarius"
 /sub_species="Somali breed"
 /db_xref="taxon:9838"
 /tissue_type="lactating mammary gland"
 CDS 25..606
 /codon_start=1
 /product="peptidoglycan recognition protein"
 /protein_id="CAC19553.1"
 /db_xref="GI:11990124"
 /translation="MTRRQVLLWALLALLSLGAARDPPACGSIYRREWRALASEC
 REHLTRPVRYVYVSHSGSHCDPPACAOAQNQSYHRLGWCVDYVNFLLGEGIL
 VYEGRWNIKGAHGPYPMNLSISIFSGNVMNPPPALRAAQNLLACVALGALR
 SNVEKGHRDVPFLSPGDRLEYIQTWSHRA"
 sig_peptide 25..87
 mat_peptide 88..603
 /product="peptidoglycan recognition protein"
 poly_a_signal 678..683
 BASE COUNT 130 a 247 c 201 g 122 t
 ORIGIN
 Query Match 54.88; Score 393.6; DB 4; Length 700;

Best Local Similarity 77.1%; Pred. No. 1,3e-77;
 Matches 507; Conservative 0; Mismatches 144; Indels 7; Gaps 2;
 QY 51 CTATGTCCGCGCGCTTATGCTGCTTGTGGGCTCTCCACCTCTTGCATCGAG 110
 Db 23 CCATGACCGGCGACTCGTGTCTCTGTGGCTCTCCCTCGGCTCTCAGCTCGAG 82
 QY 111 CGGTCAGAGACAGAAAGACCCGGCTGTGCAAGCCCATAGTCCCGGAGAGTGA 170
 Db 83 CGGCTG-----AGAAAGACCCGCGGCTGTGCAAGCCCATAGTCCCGGAGAGTGA 136
 QY 171 AGGCCCTGCAATCAGATGAGTGGCCGACACACCTGAGCCCTTACGCTATGTGGTAT 230
 Db 137 GGGCCCTGGCGCTTCGATGTCAGAGAAAGCTTAACAGCCGCGTCTACGTGTGGT 196
 QY 231 CGCACAGGCGGGGACAGCACTGCAACACCCCGCTGTGCCAGCAGAGCCCGATG 290
 Db 197 CGCACACTGCGGGACAGCACTGCGACACCCCGGCTTGTGTGGCGCAGAGCCAGAG 256
 QY 291 TGCAGACATACCAATGAAAGACACTGGGCTGTGCGACAGTGGGCTTCACTTCTGAT 350
 Db 257 TGCAGACATACCAATGAAAGACACTGGGCTGTGCGACAGTGGGCTTCACTTCTGAT 316
 QY 351 GAGAGACGGGCTGTGATACGAGAGGCGGTGGCTGGAATTCACGGGTCCTCAGGTC 410
 Db 317 GAGAGATGGGCTGTGATACGAGAGGCGGTGGCTGGAATTCACGGGTCCTCAGGTC 376
 QY 411 ACTTATGAAACCCCATGCTCATTGGCATGAGTTCATGAGGCAACTATGATGGGTC 470
 Db 377 CCACCTGGAACCCCATATGATGAGTTCCTTATGAGGCAACTATGATGAGTGC 436
 QY 471 CCACACCCAGGCGCAATCGGCGAGCCAGGCTTACTGGCTTGGGTGTGCGTACAGGAG 530
 Db 437 CCGCGGCGCGGCGCTCGGCGGACCCCAAACTGTGTGCTTGTGTGCTTGGCTGAG 496
 QY 531 CCGTGAAGTCACTATGTGCTCAAAAGACACCGGATGTGACGCTACACTCTCCAG 590
 Db 497 CCGTGAAGTCACTATGTGCTCAAAAGACACCGGATGTGACGCTACACTCTCCAG 556
 QY 591 GCAACAGCTCTACCACTATCAGATTCAGAAATTTGGCCACACTACCGCTCCCTGAGGCGT 650
 Db 557 GTACCGGCTCTACCAAAATTCAGACTGTGTACACTACACGCG-CCATGAGGCGCTTC 615
 QY 651 GTGATCCGACCCCAATTCCTCCCTCCCATGGCCAAAACCCCATCTCTCTCC 708
 Db 616 CGGCTGGCACACCGCTCCATCCCACTCTGTCAAGAAACCCCATCTCTCC 673

RESULT 4
 AY083309 688 bp mRNA linear NAM 29-MAY-2002
 LOCUS Bos taurus oligosaccharide-binding protein mRNA, complete cds.
 DEFINITION AY083309
 VERSION AY083309.1 GI:19550241
 KEYWORDS
 SOURCE Bos taurus.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 688)
 AUTHORS Tytell,C.C., Yount,N., Tran,D., Yuan,J. and Seilsted,M.E.
 TITLE Isolation, characterization, and antimicrobial properties of bovine
 oligosaccharide-binding protein. A microbicidal granule protein of
 eosinophils and neutrophils
 JOURNAL J. Biol. Chem. 277 (22), 19658-19664 (2002)
 MEDLINE 11880375
 PUBMED 22028028
 REFERENCE 2 (bases 1 to 688)
 AUTHORS Yount,N.Y., Yuan,J., Tytell,C.C. and Seilsted,M.E.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAY-2002) Pathology, UC Irvine, 440D Medical Sciences
 1, Irvine, CA 92697, USA

SOURCE Mus musculus.
ORGANISM Mus musculus
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Slayton, W.B., Rigaa, A., Hancock, J.D., Zaugg, J.R., Le, T.V.,
TITLE Treutman, M.S., Spangrude, G.J., Carroll, W.L. and Schibler, K.R.
JOURNAL Granulocyte-colony stimulating factor up-regulates expression of
REFERENCE 2 (bases 1 to 669)
AUTHORS Slayton, W.B., Rigaa, A., Hancock, J.D., Zaugg, J.R., Le, T.V.,
TITLE Treutman, M.S., Spangrude, G.J., Carroll, W.L. and Schibler, K.R.
JOURNAL Direct Submission
FEATURES Submitted (11-OCT-1999) Pediatrics, University of Utah, 50 North
source Medical Drive, Room 2A126, Salt Lake City, UT 84132, USA
Location/Qualifiers
1. 669
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="32D cl3"
/cell_type="myeloblast"
25. 573
/codon_start=1
/product="TAG7-like protein"
/protein_id="AA06335.1"
/db_xref="GI:6273361"
/translation="MLPACALALALGATSCSFIVPRSEMRALPSESSRLGHPVRY
VISHRTAGSFNSPDSCEQOARVVOHYHKNELGMCVDAVNFILGEGHYEGRGNMKG
DHRTGPIWPMISGITGFMGNFMDRVPAKRALRALNLECGVSRGFLRSNYEVKGRHDY
OSTLSPDQLYOVIOSEHYRE"
BASE COUNT 144 a 212 c 173 g 140 t
ORIGIN

Query Match 44.1%; Score 316.6; DB 10; Length 669;
Best Local Similarity 77.2%; Pred. No. 1.9e-60;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

135 CCTGTGAGAGCCCATGATGCGCCGAGAGAGTGAAGCCCTGCGATCAGAGTGGCCC 194
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db CCTCTGAGATTTCATCTGCGCCGAGAGTGAAGCCCTGCGATCAGAGTGGCCC 127
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 195 AGACCTGAGAGTGGCCCTTACGCTATGTGTGATCCACAGGGGAGGAGACAGTGA 254
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 128 GCGGCTGGGGGACAGGATGCGTACGTGTGATCTACACAGCCGAGAGCTTCTGCA 187
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 255 ACACCCCGGCTGCTGCCAGCAGAGCGCCGAGATGTGCAGCATCAACATGAAGAC 314
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 188 ACACCCCGGACTCTGTGACAGAGCGCCGAGATGTGCAGCATCAACATGAAGAC 247
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 315 TGGGCTGTGAGAGTGGGCTGACAACTCCGATGTGAGAGAGCGGCTCGTATACGAG 374
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 248 TGGGCTGTGAGAGTGGGCTGACAACTCCGATGTGAGAGAGCGGCTCGTATACGAG 307
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 375 GCGGTGCTGGAATTCAGGAGTGGCCGAGTCACTTATGAAACCCATGTCATTG 434
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 308 GCGGAGCTGGAATTCAGGAGTGGCCGAGTCACTTATGAAACCCATGTCATTG 367
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 435 GCATCAGCTTATGAGGCAATCATGATGGGTGCCACACGCCAGCCATCGGGAG 494
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 368 GCATCAGCTTATGAGGCAATCATGATGGGTGCCACACGCCAGCCATCGGGAG 427
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 495 CCCAGGCTACTGAGCCGCGGTGGGTGAGGAGCCCTGAGGTCGAATGAGTGC 554
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 428 CCCAAATCTTCTGGAATGTGGGTGTCTCGGGCTTCTGATGATCAACTATGAAGTCA 487
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 555 AAGGACACCGGAGTGTGAGGAGTCACTCTCCAGGCAACGAGTCAACACTCATCC 614
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 488 AAGGACACCGGAGTGTGAGGAGTCACTCTCCAGGCAACGAGTCAACACTCATCC 547
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 615 AGATTGGCCCACTACCG 633
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 548 AAAGCTGGAGACACTACCG 566
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

RESULT 7
AF076482 680 bp mRNA linear ROD 15-AUG-1998
LOCUS AF076482
DEFINITION Mus musculus peptidoglycan recognition protein precursor (Pgip)
ACCESSION AF076482
VERSION AF076482.1 GI:3342530
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Eukarya; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 680)
JOURNAL Kang, D., Liu, G., Lundstrom, A., Gellus, E. and Steiner, H.
AUTHORS Kang, D., Liu, G., Lundstrom, A., Gellus, E. and Steiner, H.
TITLE A peptidoglycan recognition protein in innate immunity conserved
from insects to humans
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
MEDLINE 98374308
REFERENCE 9707603
PUBMED 2 (bases 1 to 680)
AUTHORS Kang, D., Liu, G., Lundstrom, A., Gellus, E. and Steiner, H.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University,
Stockholm S-106 91, Sweden
Location/Qualifiers
1. 680
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="spleen"
1. 680
/gene="Pgip"
36. 584
/gene="Pgip"
/note="PGRP"
/codon_start=1
/product="peptidoglycan recognition protein precursor"
/protein_id="AAC31821.1"
/db_xref="GI:3342531"
/translation="MLPACALALALGATSCSFIVPRSEMRALPSESSRLGHPVRY
VISHRTAGSFNSPDSCEQOARVVOHYHKNELGMCVDAVNFILGEGHYEGRGNMKG
DHRTGPIWPMISGITGFMGNFMDRVPAKRALRALNLECGVSRGFLRSNYEVKGRHDY
OSTLSPDQLYOVIOSEHYRE"
BASE COUNT 146 a 216 c 174 g 144 t
ORIGIN

Query Match 44.1%; Score 316.6; DB 10; Length 680;
Best Local Similarity 77.2%; Pred. No. 1.9e-60;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

135 CCTGTGAGAGCCCATGATGCGCCGAGAGAGTGAAGCCCTGCGATCAGAGTGGCCC 194
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 79 CCTCTGAGATTTCATCTGCGCCGAGAGAGTGAAGCCCTGCGATCAGAGTGGCCC 138
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 135 GCGGTGCTGGAATTCAGGAGTGGCCGAGTCACTTATGAAACCCATGTCATTG 434
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 79 CCTCTGAGATTTCATCTGCGCCGAGAGAGTGAAGCCCTGCGATCAGAGTGGCCC 138
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 195 AGCAGTGAAGTGGGCTGACAACTCCGATGTGAGAGAGCGGCTCGTATACGAG 254
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 139 GCGGCTGTGAGAGTGGGCTGACAACTCCGATGTGAGAGAGCGGCTCGTATACGAG 198
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 255 ACACCCCGGCTGCTGCCAGCAGAGCGCCGAGATGTGCAGCATCAACATGAAGAC 314
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 199 ACAGCCCGGACTCGTGAACAGAGCGCCGAGATGTGCAGCATCAACATGAAGAC 258
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 315 TGGGCTGTGAGAGTGGGCTGACAACTCCGATGTGAGAGAGCGGCTCGTATACGAG 374
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 259 TGGGCTGTGAGAGTGGGCTGACAACTCCGATGTGAGAGAGCGGCTCGTATACGAG 318
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 375 GCGGTGCTGGAATTCAGGAGTGGCCGAGTCACTTATGAAACCCATGTCATTG 434
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 319 GCCAGGCTGGAACATCAAGGTGACACAGAGGCCCATCTGGAATCCCATCTATTG 378
 QY 435 GCATCAGCTTATGGGCACTACATGATCGGGTCCCAACACCCAGCCATCCGGGAG 494
 Db 379 GCATCAGCTTATGGGCACTACATGATCGGGTCCCAACACCCAGCCATCCGGGAG 438
 QY 495 CCCAGGCTCTACTGGCCCTGCGGTGTGCTCAGAGGAGCCCTGAGTCCATCTGTCTCA 554
 Db 439 CCCAATCTCTTGTGGAATGTGGGGTGTCTCGGGCTCTCTGAGATCCATGAACTCA 498
 QY 555 AAGCAGCCGGGATGTGAGCGGTACACTCTCTCCAGGCAACCGAGCTTACCACTATCC 614
 Db 499 AAGCAGCCGGGATGTGAGCGGTACACTCTCTCCAGGATGACCAACTCTATCAGTTCATCC 558
 QY 615 AGAATGGCCCACTACCG 633
 Db 559 AAGCTGGGAACTACCG 577

RESULT 8
 LOCUS BC005582 713 bp mRNA linear ROD 07-AUG-2002
 DEFINITION Mus musculus, peptidoglycan recognition protein, clone MGC:11430
 ACCESSION BC005582 IMAGE:3969014, mRNA, complete cds.
 VERSION BC005582.1 GI:13542755
 KEYWORDS MGC.
 SOURCE Mus mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Strausberg, R.

REFERENCE
 AUTHORS Submitted (02-APR-2001) National Institutes of Health, Mammalian
 TITLE Gene Collection (MGC), Cancer Genomics Office, National Cancer
 JOURNAL Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amgdbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kovis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 17 Row: J Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6679292.

FEATURES

source
 1. 713
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="MGC:11430 IMAGE:3969014"
 /tissue_type="Mammary tumor. Metallothionein-1G alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_1lb="NCI_CGAP_Mam1"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 47. 595
 /codon_start=-1
 /product="peptidoglycan recognition protein"

Query Match 43.9% Score 315; DB 10; Length 713;
 Best Local Similarity 77.0%; Pred. NO. 4.4e-60;
 Matches 384; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
 BASE COUNT 172 a 218 c 178 g 145 t
 ORIGIN

/protein_id="AA05582.1"
 /db_xref="GI:13542755"
 /db_xref="LocusID:21946"
 /translation="MELFACALIALILATSCSFYRSEMPALPSESSRLGHPYRY
 VISHRAGSPGSPSPSCDQARNVCHHNEIGKCDVAFNFIIGDGHYEGGNWIKG
 DHTGPIPNPMSIGITFMGNFMFDRPAKRALNALNLBCEVSRGLRSNYEVKGRHDV
 OSTLSPGQILQVIOSEWBYRE"
 QY 135 CCTGCTGAGCCCATAGTATGCCCCGGAACGAGTGAAGGCCCTGGCATGAGTGGCCCC 194
 Db 90 CCTCTGAGTTTCATGCTGCGCCGACAGATGATGAGGGCCCTGGCATGCGAGTCTTA 149
 QY 195 AGCAGCTAGCCTGCCCTTACGCTATGTGTGTGATTCGACACAGCGCGGACAGCTGCA 254
 Db 150 GCCGCTGGGGCACCCAGTTGCTACGATGATCTCACACACAGCCGACAGCTTCTGCA 209
 QY 255 ACACCCCGCTCTGTCGACAGCAGAGCCCGGATGTGAGCACTACCAATGAGAGAC 314
 Db 210 ACAGCCCGGACTCTCTGTGAACAGAGCCCGCAATGTGACCATTTACCAACAAGTAC 269
 QY 315 TGGGCTGTGAGCAGCTGAGGCTACCACTTCTGATTTGAGAGAGAGCGGCTGTATACGAG 374
 Db 270 TGGGCTGTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 329
 QY 375 GCCGCTGTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 434
 Db 330 GCCGAGGCTGAACTCAAGGATGACACAGAGCCCATCTGGAATCCATGATGATGATGAT 389
 QY 435 GCATCAGCTTATGGGCACTACATGATGATGATGATGATGATGATGATGATGATGATGAT 494
 Db 390 GCATCAGCTTATGGGCACTACATGATGATGATGATGATGATGATGATGATGATGATGAT 449
 QY 495 CCCAGGCTCTACTGGCCCTGCGGTGTGCTCAGAGGAGCCCTGAGTCCATCTGTCTCA 554
 Db 450 CCCAATCTCTTGTGGAATGTGGGGTGTCTCGGGCTCTCTGAGATCCATGAACTCA 509
 QY 555 AAGCAGCCGGGATGTGAGCGGTACACTCTCTCCAGGCAACCGAGCTTACCACTATCC 614
 Db 510 AAGCAGCCGGGATGTGAGCGGTACACTCTCTCCAGGATGACCAACTCTATCAGTTCATCC 569
 QY 615 AGAATGGCCCACTACCG 633
 Db 570 AAGCTGGGAACTACCG 588

RESULT 9
 LOCUS MNRNATMS1 678 bp mRNA linear ROD 30-JUL-1998
 DEFINITION M.musculus mRNA for TAG7 protein.
 ACCESSION X86374.1 GI:887524
 VERSION X86374.1 GI:887524
 KEYWORDS tms1 gene.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Kustlikova, O.S., Kiselev, S.L., Borodulina, O.R., Senin, V.M.,
 Afanas'eva, A.V. and Kabishev, A.A.
 Cloning of the tag7 gene expressed in metastatic mouse tumors
 Genetika 32 (5), 621-628 (1996)
 MEDLINE 96321104
 PUBMED 8755036
 REFERENCE 2 (bases 1 to 678)
 AUTHORS Kustlikova, O.S.
 DIRECT SUBMISSION Submitted (18-APR-1995) O.S. Kustlikova, Inst of Gene Biology
 JOURNAL

REFERENCE 3 (bases 1 to 678)
 AUTHORS Kiselev S.L., Kustikova O.S., Korobko E.V., Prokhorchuk E.B., Kabishov A.A., Lukanidin E.M., and Georgiev G.P.
 TITLE Molecular cloning and characterization of the mouse tag7 gene encoding a novel cytokine
 JOURNAL J. Biol. Chem. 273 (29), 18633-18639 (1998)
 MEDLINE 98325081
 PUBMED 9660837
 COMMENT On Jul 6, 1995 this sequence version replaced gi:785010.
 FEATURES
 source Location/Qualifiers
 1. 678
 /organism="Mus musculus"
 /strain="A/Sn"
 /db_xref="taxon:10090"
 /tissue_type="adenoecarcinoma"
 /dev_stage="adult"
 36..491
 /gene="tag7"
 36..491
 /gene="tag7"
 /codon_start=1
 /protein_id="CAA60133.1"
 /db_xref="GI:887525"
 /db_xref="MGI:107415"
 /db_xref="SPTREMBL:O62185"
 /translation="MLFACILALILGATSCFIVPRSRNALPSCSRGLCHPVRVY
 VISHRAGSRNPSDSCEDQARVQVYHKNELSCDVAINFLIGEGHYHEGGMWIKG
 DHGPIRPMNSGITTFMGFMKVRKAGPCCPKSSGIWVSGLEPIOL"

BASE COUNT 146 a 216 c 173 g 143 t
 ORIGIN

Query Match 42.1%; Score 302.6; DB 10; Length 678;
 Best Local Similarity 76.8%; Pred. No. 2.6e-57;
 Matches 383; Conservative 0; Mismatches 114; Indels 2; Gaps 1;

QY 135 CCTGCTGACGCCCATATGATGCCCCGGAGACGAGTGAAGGCCCTGCGATGAGTGGCCCC 194
 DB 79 CCTCTGCGAGTTTCATCGCGCCCGAGTGAAGGCCCTGCGATGAGTGGCCCC 138
 QY 195 AGCAGCTGAGCCTGCGCCTTACGCTATGCTGATGCTGACGACGCGGCGAGCAGCTGCA 254
 DB 139 GCCGCTGGGGGACCCAGTGGCTGCTGATGCTGACGACGCGGCGAGCTTCTGCA 198
 QY 255 ACACCCCGCCCTGCTGACGACGAGGCCCGGAGATGTCAGCAGTACCATGAAGACAC 314
 DB 199 ACAGCCCGGAGCTCTGTGAACAGAGGCCCGGAGATGTCAGCAGTACCATGAAGAC 258
 QY 315 TGGGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 374
 DB 259 TGGGCTGCTGAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
 QY 375 GCCGCTGCTGAGTACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 434
 DB 319 GCCGAGGCTGAGACATCAAGGCTGACACAGGCCCATCTGGAATCCATGCTATTG 378
 QY 435 GCATGAGCTTCAATGAGGACATGATGAGTGGGTCGCCACACCCAGGCCATCCGGGAG 494
 DB 379 GCATGAGCTTCAATGAGGACATGATGAGTGGGTCGCCACACCCAGGCCATCCGGGAG 436
 QY 495 CCCAGGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 554
 DB 437 CCCRAATCTCTGGAATCTGGGGTGTCTCGGGGCTTCTGATGATCAATGAAAGTCA 496
 QY 555 AAGGACACCGGAGTGTGACAGCTCTCTCCAGGACACAGCTCTACACCTATCC 614
 DB 497 AAGGACACCGGAGTGTGACAGCTCTCTCCAGGACACAGCTCTACACCTATCC 556
 QY 615 AGAATGGCCACACTACCG 633
 DB 557 AAAGCTGGGAACACTACCG 575

RESULT 10
 AC007785/c
 LOCUS
 DEFINITION Homo sapiens chromosome 19, BAC 282485 (CIT-B-344H19), complete
 ACCESSION AC007785
 VERSION AC007785.1
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 166500)
 Lamerdin, J.E., McCready, P.M., Richardson, P., Sakladis, G.,
 Burkhardt-Schultz, R., Gordon, L., Scott, D., Johnson, G., Stillwagen, S.,
 Phan, H., Velasco, N., Do, L., Regala, W., Terry, A., Dangnan, L.,
 Erlar, A., Christensen, M., Georges, A., Avila, J., Altix, C.,
 Andreise, T., Amico-Keller, G., Coefield, J., Duarte, S., Lucas, S.,
 Bruce, R., Thomas, P., Quan, G., Krommiller, B., Arellano, A.,
 Sanders, C., Ow, D., Nolan, M., Trong, S., Kobayashi, A., Olsen, A.S. and
 Carrano, A.V.
 Sequence analysis of a 2.5 Mb region in 19q13.2-13.3 between APOE
 and C5R1
 2 (bases 1 to 166500)
 Lamerdin, J.E.
 Direct Submission
 Submitted (11-JUN-1999) Joint Genome Institute, Lawrence Livermore
 National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
 Map and sequence oriented from 9 centimorgans to telomere, BAC 282485
 overlaps cosmid R30477 to the left from bases 1 to 4,419 of this
 accession, and overlaps BAC 264576 (CIT-B-297N14; AC006262) from
 bases 146,203 to 166,500. Additional chromosome 19 map and sequence
 information may be obtained at:
 http://www.bio.livnl.gov/db/brp/genome/genome.html.
 Location/Qualifiers
 1. 166500
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19q13.3 between DM and C5R1"
 /clone="CIT-B-344H19"
 /clone_11b="Cal Tech CIT-B BAC 11library"
 /note="LNL clone name: BC282485"
 70..145
 /rpt_family="TGAAn"
 complement(177..453)
 /rpt_family="Alubp"
 complement(472..712)
 /rpt_family="MIR"
 john<746..904,1157..1239,4294..4451,5875..6036,
 12784..12994,13080..13162,13903..13958,14053..14128,
 14429..14568,14726..14804,14982..15039,15813..15865,
 16025..16122)

FEATURES

source

repeat_region

repeat_region

repeat_region

CDS

/note="Hypothetical arginine-rich gene product;
 Hypothetical CDS constructed from overlapping ESTs and
 X-ray predictions. EST coverage is lacking for some
 portions of the CDS. Gene may be alternatively spliced, as
 multiple transcripts map into this region; alternatively,
 these individual clones may represent artifactual events
 during RT and cDNA creation"
 /codon_start=3
 /evidence-not_experimental
 /product="BC282485_1"
 /protein_id="AAB38244.1"
 /db_xref="GI:5042405"
 /translation="TLADDPAGLQVDFVGVAVVWVSGOYLEVEVDRMADO
 WRGEFAGFIEDLTHKNGENKOFNIPCMESALQSSSEYTLDTTDESLENRK
 MGRGSLAPRSALQNSKRYILLYSEFDEIRHPLPLPYGKDPVAVLQITRLKE
 ELGRLQIDGONTDTRENIYHLEQVSRILASKRELEADLGRREVALAGRAOE
 AELRGLVRELELROREGLHVRVAGRGDCCRLAKELEAKSESLNRLTLT
 SELALYRKGGRSGRGRPARPSPPTGGRALRDPPLAFVAKERKKEIQMKQOORN

```

RUGSGSGDGPBSVSNROTOPPALTRGADAPNRNRSSSVDSFRSSASCSCL
EDFESLSRGGHRRGKPPSPMGSGNMKSPVRSRHQKSLANSNGWVPIKESSE
HQAADMAIDARIKALQEHYMNRLDMRS"
746. .904
/misc_feature
/note-"DSS similarity to overlapping ESTs:
N39722 yx92e04.r1 Homo sapiens CDNA clone 269214 5';
(12. .170); 99% identity. -AA283147 zt14d07.r1 NCI CGAP GCBI
Homo sapiens CDNA clone IMAGE:713101 5'; (6. .163); 99%
identity. -AA143160 z149d07.r1 Soares pregnant uterus NBHPU
Homo sapiens CDNA clone 505261 5'; (1. .43); 100%
identity. -AA142881 z149d07.s1 Soares pregnant uterus NBHPU
Homo sapiens CDNA clone 505261 3'; (570. .527); 64%
identity."
1157. .1239
/misc_feature
/note-"Predicted exon, program: gra112exons_human_1.3,
frame: 0, quality: good, score: 62.000-DSS similarity to
overlapping ESTs:
N39722 yx92e04.r1 Homo sapiens CDNA clone 269214 5';
(171. .255); 95% identity. -AA283147 zt14d07.r1
NCI CGAP GCBI Homo sapiens CDNA
clone IMAGE:713101 5'; (164. .247); 92% identity. -AA143160
z149d07.r1 Soares pregnant uterus NBHPU Homo sapiens CDNA
clone 505261 5'; (44. .126); 100% identity. -N28707
yx66d11.r1 Homo sapiens CDNA clone 266709 5'; (1. .66); 98%
identity. -AA142881 z149d07.s1 Soares pregnant uterus NBHPU
Homo sapiens CDNA clone 505261 3'; (526. .438); 82%
identity."
1536. .1634
/repeat_region
/rpt_family-"CA)n"
complement(1673. .1973)
/repeat_region
/rpt_family-"Alusx"
complement(2040. .2185)
/repeat_region
/rpt_family-"LINE2"
complement(2200. .2498)
/repeat_region
/rpt_family-"Alusx"
complement(2576. .2746)
/repeat_region
/rpt_family-"Alusg/x"
complement(2747. .3044)
/repeat_region
/rpt_family-"Alusx"
3048. .3750
/repeat_region
/rpt_family-"MER64"
3841. .4060
/repeat_region
/rpt_family-"MER64"
4294. .4451
/misc_feature
/note-"DSS similarity to overlapping ESTs:
N39722 yx92e04.r1 Homo sapiens CDNA clone 269214 5';
(256. .438); 81% identity. -AA283147 zt14d07.r1
NCI CGAP GCBI Homo sapiens CDNA
clone IMAGE:713101 5'; (248. .400); 99% identity. -AA143160
z149d07.r1 Soares pregnant uterus NBHPU Homo sapiens CDNA
clone 505261 5'; (127. .281); 97% identity. -N28707
yx66d11.r1 Homo sapiens CDNA clone 266709 5'; (67. .222);
98% identity. -AA142881 z149d07.s1 Soares pregnant uterus
NBHPU Homo sapiens CDNA clone 505261 3'; (437. .275); 97%
identity."
5875. .5898
/repeat_region
complement(4789. .5083)
/rpt_family-"Alusx"
5875. .5972
/misc_feature
/note-"DSS similarity to overlapping ESTs:
N39722 yx92e04.r1 Homo sapiens CDNA clone 269214 5';
(439. .506); 71% identity. -AA143160 z149d07.r1 Soares
pregnant uterus NBHPU Homo sapiens CDNA clone 505261 5';
(282. .379); 100% identity. -AA142881 z149d07.s1 Soares
pregnant uterus NBHPU Homo sapiens CDNA clone 505261 3';
(274. .175); 95% identity."
5875. .5898
/misc_feature
/note-"DSS similarity to N28707 yx66d11.r1 Homo sapiens
CDNA clone 266709 5'; (223. .246); 100% identity."
5906. .6036
/misc_feature
/note-"Predicted exon, program: gra112exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
complement(6148. .6443)
/rpt_family-"Alusx"

```

```

repeat_region complement(6608. .6648)
/rpt_family-"LINE2"
repeat_region complement(7118. .7340)
/rpt_family-"MIR"
repeat_region complement(7538. .7639)
/rpt_family-"CA)n"
repeat_region complement(7792. .8054)
/rpt_family-"MIR"
repeat_region complement(8789. .9083)
/rpt_family-"AluJo"
9456. .9520
repeat_region
/rpt_family-"MER5B"
complement(9539. .9841)
/rpt_family-"Alusx"
9909. .9944
repeat_region
/rpt_family-"MER5B"
10099. .10401
repeat_region
/rpt_family-"Alusx"
complement(10587. .10752)
/rpt_family-"AluJ"
complement(10753. .11057)
/rpt_family-"Alusx"
complement(11065. .11191)
/rpt_family-"AluJ"
11322. .11619
repeat_region
/rpt_family-"Alusg"
11775. .11814
repeat_region
/rpt_family-"AT_rich"
complement(11864. .11885)
/rpt_family-"AT_rich"
complement(11914. .11976)
/rpt_family-"Alus"
complement(12055. .12088)
/rpt_family-"AT_rich"
complement(12186. .12327)
/rpt_family-"LINE2"
complement(12334. .12384)
12784. .12994
/misc_feature
/note-"Predicted exon, program: gra112exons_human_1.3,
frame: 1, quality: excellent, score: 92.000"
12847. .12908
repeat_region
/rpt_family-"GC_rich"
12895. .12994
/misc_feature
/note-"DSS similarity to AA765166 nz79c05.s1 NCI CGAP GCBI
Homo sapiens CDNA clone IMAGE:1301672; (222. .123); 97%
identity."
13002. .13032
repeat_region
/rpt_family-"GC_rich"
13080. .13162
/misc_feature
/note-"Predicted exon, program: gra112exons_human_1.3,
frame: 2, quality: excellent, score: 100.000-DSS
similarity to AA765166 nz79c05.s1 NCI CGAP GCBI Homo
sapiens CDNA clone IMAGE:1301672; (122. .48); 100%
identity."
complement(13799. .13954)
/rpt_family-"GC_rich"
20681
Query Match 41.6%; Score 299; DB 9; Length 166500;
Best Local Similarity 92.6%; Pred. No. 8.3e-57;
Matches 314; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
OY 8 ACGGCGCCGACAGAGCGTGGGCGCCGCGACATGATCTCAGCCACTATGTCGCCGCTCT 67
DB 20801 ACCGGGCGCTCCAGCGGTCTCCGCGACCTCGCCCTGCATATGTCGCCGCTCT 20742
OY 68 ATGCTGCTGCTGCGGCTCTCCCGAGCTCTGACTCGAGCGGCTCAGAGACAGAA 127
DB 20741 ATGCTGCTGCTGCGGCTCTCCCGAGCTCTGACTCGAGCGGCTCAGAGACAGAA 20682
OY 128 GACCCGCGCTGCTGACAGCCCAATAGTCCCGGACGAGTGAAGGCCCTGCTCAGAG 187
DB 20681 GACCCGCGCTGCTGACAGCCCAATAGTCCCGGACGAGTGAAGGCCCTGCTCAGAG 20622

```

	Query Match	Similarity	41.2%	Score 295.6	DB 10	Length 630
	Best Local	Similarity	74.3%	Pred. No. 9,4e-56		
	Matches 373	Conservative	0	Mismatches 129	Indels 0	Gaps 0
QY	132	CGGCGTGTGACAGCCCATAGTGCCTCCGGAACAGAGGAAAGCCCTGGCATCGAGATGCCG	191			
DB	63	CAGACTCTGTGCTTTTGTGTCGTCCCGCCGAGTAGTGGAAAGGCCCTGGCCATCCGATGTGCT	122			
QY	192	CCCAGACTGAGCCTGCCCTTACGCTATGTGGTATCGACACAGGGGGGAGCAGCT	251			
DB	123	CCAAAGGCTGGAAGAAACCAAGTCGCGCTAGCTGTGTATCATACACAGCGCGGAGCTTCT	182			
QY	252	GCACACGCCCGCGCTGTGCCAGCAGACAGCCCGGAATGTGACGACACTACACATTAAGA	311			
DB	183	GCAGAGCCCCAGACTCTCTGAAACACAGAGCCCGCATGTGACAGCTTACCAATTAAG	242			

CDS	<2..>259	/gene="pglyrp"	/function="bacterial binding, anti-inflammatory, cytokine"
		/codon_start=1	/product="peptidoglycan recognition protein"
		/protein_id="CAC83647.1"	/db_xref="GI:17402405"
		/translation="AGSTCDNPASCRQVONQHYHVPRLHWCDDGYNFLIGEDGLV	
		ECRGNTVGAHSGPWNPLSLGISIFMGNTMNNVPPARAIRAA"	
		<2..>259	
mat_peptide		/gene="pglyrp"	/product="peptidoglycan recognition protein"
BASE COUNT	47 a	86 c	83 g
ORIGIN			43 t

Matches 213; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 238 GGGGGGAGCAGCTGACACACCCCGCTCTGTCAGCAGCGCCGGAATGTGCACA 297
 Db 1 GGGGGGAGCAGCTGACACACCCCGCTCTGTCAGCAGCGCCGGAATGTGCACA 60

QY 298 CTACACATGAAAGACTGGCTGTGACAGCTGAGCTTACACTTCCGATTTGGAGAGA 357
 Db 61 CTACACAGTGGAGACGCTGACAGCTGTGACAGCTGAGCTTACACTTCCGATTTGGAGAGA 120

QY 358 CGGGCTCTATACAGAGCGCTGCTGAGAACTTACAGGGTGCCCACTAGTCACTTATG 417
 Db 121 CGGGCTCTATATGATGAAGCGCGGGCTGGAACACCGTGGCGCCCACTAGTCCACAGT 180

QY 418 GAACCCCATGTCATTTGTCATCAGCTTATGAGCACTACATGATCGGGTCCCAACACC 477
 Db 181 GAACCCCTGTCTGTGGGCACTCTCTTCATGGGCAACTACATGATCGGGTCCCAACACC 240

QY 478 CCAGGCCATCCGGCGCAGCC 496
 Db 241 CCGTGCCATGAGGGCGGCC 259

RESULT 13
 CDR409286 5358 bp DNA linear NAM 19-DEC-2001
 LOCUS Camelus dromedarius pgrp gene for peptidoglycan recognition
 ACCESSION AJ409286
 VERSION AJ409286.1 GI:17939892
 KEYWORDS peptidoglycan recognition protein; pgrp gene.
 SOURCE Arabian camel.
 ORGANISM Camelus dromedarius
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.

REFERENCE 1
 AUTHORS Kappeler S.R., Parah Z. and Puhon Z.
 TITLE The peptidoglycan recognition protein is expressed in the lactating mammary gland of camels and binds to lactic acid bacteria
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 5358)
 AUTHORS Kappeler S.R.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAR-2001) Kappeler S.R., Food Science, Swiss Federal Institute of Technology, Zurich, LFO F26, ETHZ, CH-8092 Zurich, SWITZERLAND

FEATURES
 source location/Qualifiers
 1..5358 /organism="Camelus dromedarius"
 /db_xref="taxon:9838"
 /cell_type="lymphocyte"
 /country="United Arab Emirates"
 /note="breed Arabian"
 repeat_region 44..254 /rpt_family="SINE/MIR"
 protein_bind 203..210 /bound_moiety="STAT5"
 /function="stage- and tissue-specific expression of milk proteins"
 repeat_region 478..545 /note="similar AF217601"
 /rpt_family="tRNA-Ala-CCG"
 898..1063 /rpt_family="SINE/MIR"
 gene 1585..5358 /gene="pgrp"
 TATA_signal 1585..1590 /gene="pgrp"
 mRNA join(1599..1921,4528..4649,5086..5358)
 exon 1599..1921 /gene="pgrp"
 /number=1

5'UTR 1599..1640 /gene="pgrp"
 CDS /evidence="experimental
 join(1641..1921,4528..4649,5086..5264)
 /gene="pgrp"
 /function="binding to bacterial peptidoglycan (low peptide cross-linked and low modified)"
 /codon_start=1
 /product="peptidoglycan recognition protein"
 /protein_id="CAC84130.1"
 /db_xref="GI:17939893"
 /translation="MTRRCVLLVALLLALLSLGAAREDPACGCTVPRRERALLSEC
 NRELRTPRVIVSHRAGSHCDTPASCAQQAQNVQSYHVRNLGMDVGNFLIGEDGL
 VYEGGNRIKGAHAGPTWNPISIGISFMGNVNRVPPPRALRAAONLLACGVALGALR
 SNYEKGRHDVQPTLSPGDRUYEIIQTWSHYRA"
 1641..1703
 sig_peptide /gene="pgrp"
 /evidence="experimental
 join(1704..1921,4528..4649,5086..5261)
 /gene="pgrp"
 /product="peptidoglycan recognition protein (PGRP)"
 /evidence="experimental
 1922..4527
 /gene="pgrp"
 /number=1
 2636..2939
 /rpt_family="LINE/L2"
 4213..4394
 /rpt_family="SINE/MIR"
 exon 4528..4649 /gene="pgrp"
 /number=2
 4650..5085
 /gene="pgrp"
 /number=2
 4884..5031
 /rpt_family="SINE/MIR"
 repeat_region 5086..5358 /gene="pgrp"
 /number=3
 5336..5341
 /gene="pgrp"
 exon 5358 /gene="pgrp"
 polyA_signal 5358 /gene="pgrp"
 polyA_site 5358 /gene="pgrp"
 BASE COUNT 1287 a 1351 c 1533 g 1187 t
 ORIGIN

Query Match 22.5%; Score 161.4; DB 4; Length 5358;
 Best Local Similarity 74.1%; Pred. No. 6.1e-26;
 Matches 220; Conservative 0; Mismatches 71; Indels 6; Gaps 1;

QY 51 CTATGTCCCGCCGCTCTATGCTGTGCTGGCTCTCCCAAGCCTCTTCGACTCGAG 110
 Db 1639 CCATGACCCCGGACTGCGCTTCTGCTGCGCTCTCTCCCTCCGCTCAGCCTGGAG 1698

QY 111 CGGGCTCAGAGAGACAACACCGGCTGCTGCAAGCCATTAAGTCCCGGAAAGAGTGA 170
 Db 1659 CGGGCTG-----AGAAAGCCCGCGGCTGCGCTTCATCTGTCGCCCGCGAGAGTGA 1752

QY 171 AGCCCTGGCATCAGATGCGCCGCGACACCTGAGCTGCTTACGCTATGTCGCTAT 230
 Db 1753 GGGCCCTGGCGTCCGAGTGCAGAGAAAGGCTTAACACAGCCGGTGCCTACGTGTGTGT 1812

QY 231 CGCACAGCGCGGGGAGCAGCTGCAACACCCCGCTGTGTCAGCAGCAGGCCCGGAATG 290
 Db 1813 CGCACACTGCGGGGAGCAGCTGCAACACCCCGCTTGTGTCGCGCAGCAGGCCCGGAATG 1872

QY 291 TGCAGCACTACACATGGAAGACACTGGCTGTGTCAGCTGAGCTGAGCTTCTCTGA 347
 Db 1873 TGCAGCACTACACATGTCGGAACCTGGCTGTGTCAGCTGAGCTGAGCTGAGTGA 1929

RESULT 14

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 12, 2002, 02:34:35 ; Search time 2689 Seconds

(without alignments)
2067.176 Million cell updates/sec

Title: US-09-462-625-4
Perfect score: 1051
Sequence: 1 MILAWALPSILRLGAAQETE.....LSPGNQYHLIONMPHYRSP 191

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL-frame+ p2n.model -DEV-rlp
-O/cgn2.1/USPTO.spoal/US0946265/runat.07112002.101525.24840/app.query.fasta.1.327
-DB-GenEmbl -GFWT-fastap -SUFFIX-rge -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human4.0.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTPMT-pto -NORW-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0946265.ecgn.1.1.3637 @runat.07112002.101525.24840 -NCPU=6 -ICPU=3
-NO.XLPYX -NO.MMAP -LARGROUPEY -NEG.SCORES=0 -WAIT -LONCLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -WAIT -LONCLOG -DEV.TIMEOUT=120
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1051	100.0	690	9 AF076483	AF076483 Homo sapi
2	1051	100.0	724	9 AF242517	AF242517 Homo sapi
3	772	73.5	700	4 CDR131676	AJ131676 Camelus d
4	717	68.2	688	4 AY083309	AY083309 Bos tauru
5	692	65.8	549	6 AR124884	AR124884 Sequence
6	692	65.8	669	10 AF193843	AF193843 Mus muscu
7	692	65.8	680	10 AF076482	AF076482 Mus muscu
8	692	65.8	713	10 BC005582	BC005582 Mus muscu
9	667.5	63.5	630	10 AF154114	AF154114 Rattus no
10	650.5	61.9	678	10 NM000001	NM000001 M. musculus
11	505.5	48.1	166500	9 AC007785	AC007785 Homo sapi
12	441	42.0	1173	9 AY035376	AY035376 Homo sapi
13	414.5	39.4	166867	3 AC007085	AC007085 Drosophila
14	414.5	39.4	169509	3 AC007303	AC007303 Drosophila
15	414.5	39.4	172838	2 AC020451	AC020451 Drosophila
16	414.5	39.4	259336	3 AE003836	AE003836 Drosophila
17	413	39.3	1110	6 AX119918	AX119918 Sequence
18	413	39.3	1128	6 AX119915	AX119915 Sequence
19	413	39.3	1194	6 AY035377	AY035377 Homo sapi
20	405.5	38.6	652	3 AF075442	AF075442 Drosophila
21	403.5	38.6	651	3 AY060645	AY060645 Drosophila
22	395.5	37.6	1858	9 AF242518	AF242518 Homo sapi
23	388.5	37.0	670	3 AF076481	AF076481 Trichoplus
24	386	36.7	259	4 SSC310355	AJ103555 Sus scrofa
25	379	36.1	1162	3 AF075441	AF075441 Drosophila
26	379	36.1	1162	3 AY075293	AY075293 Drosophila
27	369	35.1	718	3 AF441723	AF441723 Bombyx mori
28	368	35.0	5358	4 CDR409286	AF441723 Bombyx mori
29	353.5	33.6	1869	4 AF500096	AF500096 Camelus d
30	348.5	33.2	1794	3 AF384856	AF384856 Homo sapi
31	348	33.1	753	6 E16480	E16480 Bombyx mori
32	347.5	33.1	590	6 E16479	E16479 Bombyx mori
33	347.5	33.1	2751	9 AK055882	AK055882 Homo sapi
34	347	33.0	1374	3 AY058258	AY058258 Drosophila
35	347	32.9	735	3 AF313391	AF313391 Drosophila
36	346	32.9	735	3 AB016249	AB016249 Bombyx mori
37	339	32.3	622	3 AB017520	AB017520 Bombyx mori
38	339	32.3	1107	6 AX119920	AX119920 Sequence
39	339	32.3	1125	6 AX119917	AX119917 Sequence
40	338	32.2	1805	10 BC019396	BC019396 Mus muscu
41	338	32.2	159930	2 AC014935	AC014935 Drosophila
42	338	32.2	186002	3 AC006933	AC006933 Drosophila
43	338	32.2	201313	3 AC010688	AC010688 Drosophila
44	338	32.2	280887	3 AE003526	AE003526 Drosophila
45	336.5	32.0	673	3 AB017519	AB017519 Bombyx mori

RESULT 1

ALIGNMENTS

AF076483 690 bp mRNA linear PRI 15-AUG-1998
 LOCUS AF076483
 DEFINITION Homo sapiens peptidoglycan recognition protein precursor (PGRP)
 ACCESSION AF076483
 VERSION AF076483.1 GI:3342532
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 690)
 AUTHORS Kang, D., Liu, G., Lundstrom, A., Gelius, E. and Steiner, H.
 TITLE A peptidoglycan recognition protein in innate immunity conserved from insects to humans
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998) 448/8
 MEDLINE 98374308
 PUBMED 9707603
 REFERENCE 2 (bases 1 to 690)
 AUTHORS Kang, D., Liu, G., Lundstrom, A., Gelius, E. and Steiner, H.
 TITLE Direct Submission
 JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University, Stockholm S-106 91, Sweden
 FEATURES
 source location/Qualifiers
 1..690
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="bone marrow"
 1..690
 /gene="PGRP"
 /gene="PGRP"
 45..635
 /gene="PGRP"
 /codon_start=1
 /product="peptidoglycan recognition protein precursor"
 /protein_id="AAC31822.1"
 /db_xref="GI:3342532"
 /translation="MSRRSLMALWALPSLRIGAEDTEDPACCSPIVPRNEMKALAS
 ECAQHLSPLRYVVSHTAGSSCNTPASCCOQARNOYHMKTLGWCDDVYINFLIGD
 GLVYEGRGNNFTGAHSGHLNPMISIGISFWGNMDRVPPTQAIIRAAAGLLACGVAQGA
 LRSNVYLKGRDVRQLTSPGNOLYHLIONPHRSP"
 45..104
 /gene="PGRP"
 /note="putative"
 105..632
 /gene="PGRP"
 /note="putative"
 mat.peptide /product="peptidoglycan recognition protein"
 BASE COUNT 129 a 188 g 123 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,31e-93 Length: 690
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-462-625-4 (1-191) x AF076483 (1-690)
 QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGluYAlaAlaGlnGluThrGlu 20
 Db 60 ATGCTGCTTCCCTGGGCTCTCCAGCCCTCTGACTCGAGCGGCTCAGAGACAGAA 119
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpIleAlaLeuAlaSerGlu 40
 Db 120 GACCCGGCTGCTGACGCCCATAGTGCCTCCGAGACAGATGGAAGCCCTGCGCATCAGAG 179
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyrValValYalSerHisThrAlaGlySer 60
 Db 180 TGGGCCAGACCTGAGCCCTTACGCTATGTGTGATCGCACAGCGCGGCGAGC 239
 QY 61 SerCysAsnThrProAlaSerCysGlnGlnAlaAlaArgAsnValGlnHisTyrHisMet 80

Db 240 AGCTGACACACCCCGCCCTCTGCCAGACAGAGCCCGGAATGTGCACACTACACATG 299
 QY 81 LysThrLeuGlyTyrPcysAspValGlyTyrAsnPhleuLeuGlyGluAspGlyLeuVal 100
 Db 300 AAGACACTGGCTGTGGTGGACAGCTGGGCTCAACTTCTCTATGGAGAAACAGCGGCTCGTA 359
 QY 101 TyrGlnGlyArgGlyTyrPasnPherThrGlyAlaHisSerGlyHisLeuTrpAsnProMet 120
 Db 360 TACGAGGCGCGTGGTGGAACTTCAACGGGTGCCACTCGTCACTTATGGAACCCATG 419
 QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 Db 420 TCCATTGGCATAGCTTCATGGGCAACATCATGATCGGCTCCACACCCAGCGCATC 479
 QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 Db 480 CGGGACGCCCAAGGGGTCTACTGGCGTGGCTGCTGAGGAGCCCTGAGTCCACATAT 539
 QY 161 ValLeuLysGlyHisAspArgValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 Db 540 GTGCTCAAGAGACACCGGAGATGTGACGGCTACACTCTCTCCAGGACACGACTTACAC 599
 QY 181 LeuIleGlnAsnTrpProHisTyrArgSerPro 191
 Db 600 CTCATCCAGAAATGGCCACTACCGCTCCCC 632
 RESULT 2
 AF242517 724 bp mRNA linear PRI 15-AUG-2000
 LOCUS AF242517
 DEFINITION Homo sapiens hypothetical protein SBBI68 mRNA, complete cds.
 ACCESSION AF242517
 VERSION AF242517.1 GI:9802032
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 724)
 AUTHORS Wen, T., Zhang, W. and Cao, X.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAR-2000) Department of Immunology, Second Military Medical University & Shanghai Brilliance Biotechnology Institute, 800 Xiangyin Rd., Shanghai 200433, P.R. China
 FEATURES
 source location/Qualifiers
 1..724
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 30..620
 /codon_start=1
 /product="hypothetical protein SBBI68"
 /protein_id="AF99598.1"
 /db_xref="GI:9802032"
 /translation="MSRRSLMALWALPSLRIGAEDTEDPACCSPIVPRNEMKALAS
 ECAQHLSPLRYVVSHTAGSSCNTPASCCOQARNOYHMKTLGWCDDVYINFLIGD
 GLVYEGRGNNFTGAHSGHLNPMISIGISFWGNMDRVPPTQAIIRAAAGLLACGVAQGA
 LRSNVYLKGRDVRQLTSPGNOLYHLIONPHRSP"
 BASE COUNT 161 a 251 c 186 g 126 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,44e-93 Length: 724
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-462-625-4 (1-191) x AF242517 (1-724)
 QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGluYAlaAlaGlnGluThrGlu 20
 Db 45 ATGCTGCTTCCCTGGGCTCTCCAGCCCTCTGACTCGAGCGGCTCAGAGACAGAA 104

```

/organism="Camelus dromedarius"
/sub_species="Somali breed"
/db_xref="taxon:9838"
/tissue_type="lactating mammary gland"
25. 606
/codon_start=1
/product="peptidoglycan recognition protein"
/protein_id="CACI9553.1"
/db_xref="GI:11990124"
/translation="MTEHCIVLWVLLALISGARDEPPAGCSVYPRREKRLASLSE
RYERLTPRVYVYSHGSGCDTPASCADQAOAVOSYHRLNGWCDVGYNLLIEDGL
VYBERGNIGKAGHGPWVNIISGISFEMNNVRRPPRLAALQNLALGVALGALR
SNIEYKCHRDVQPLSGKGRLYLIIITWSHYA"

```

OY	1	AspProAlaCysCysSerProIleValProArgasnGluTrpLysAlaLeuAlaSerGlu	40
Db	40	GTCGTCTCCGCTCGGGCCTCTCCGCCCTCCAGCCCGAGCGGCCTCGA-----GAA	93
OY	21	AspProAlaCysCysSerProIleValProArgasnGluTrpLysAlaLeuAlaSerGlu	40
Db	94	GACCCCCCGGCTCCGCTCCATGTGCCCCCGGAGAGGAGGCCCTGCGCTCCGAG	153
OY	41	CysAlaGlnHisLeuSerLeuProLeuArgTyrValValSerHisThrAlaGlySer	60
Db	154	TGCAGCAAAAAGGCTTAACAACGGCGGTGCGTAGTGTTGGTTCGCACACTGCGGGAGC	213
OY	61	SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTyrHisMet	80
Db	214	CACGCGCACACCOCGGCTTGTCGGCCACAGCCCGACGAACCTGCMAAACCTACACATGTG	273
OY	81	LysThrLeuGlyLTPCysAspValGlyTyrAsnPhenLeuIleGlyGluAspGlyLeuVal	100
Db	274	CGGAACTCGGGCTGGTGGCGACGTGGGCTACAACTTCTCTGATCCGAGAAGATGGGCTGTG	333
OY	101	TyrGlnGlyArgGlyLTPAsnPhenThrGlyAlaHisSerGlyHisLeuThrAsnPromet	120
Db	334	TACGAAAGGCCGAGGCTGGAAACATCAAGGGGCGCCACGACGAGTCCCACCTGGANCCCCATA	393
OY	121	SerIleGlyLieserPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle	140
Db	394	TCCATPAGGATCTCTTCATTAGGGCAACTATATGAATCGATGGCCCCCGGCCGCCCTTC	453
OY	141	ArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaIleGlnGlyAlaLeuArgSerAsnTyr	160
Db	454	CGGGCAGCCCAAGAAATGCTGGCTTGTGGCTGTGGCTCGGAGCCCTGATGATCCAATCA	513
OY	161	ValLeuLysGlyHisIstArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis	180
Db	514	GAGGTCAAAGACACCGGATGTGCACGCCGACCTCTCTCCAGGTGACCGGCTCTACGANA	573
OY	181	LeuIleGlnHisThrProHisTyrArgSer	190
Db	574	ATCATCCAGACTTGGTCACTACATCAACGCCGA	603
RESULT 4			
LOCUS	AY083309	688 bp	mRNA
DEFINITION	Bos taurus oligosaccharide-binding protein mRNA, complete cds.	linear	MAM 29-MAY-2002
ACCESSION	AY083309		
VERSION	AY083309.1	GI:19550241	
KEYWORDS			
SOURCE	Bos taurus.		
ORGANISM	Bos taurus		
Eukaryota:	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia:	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
Bovidae;	Bovinae; Bos.		
REFERENCE	1 (bases 1 to 688)		
AUTHORS	Tydeli,C.C., Yount,N., Tran,D., Yuan,J. and Selsted,M.E.		
TITLE	Isolation, characterization, and antimicrobial properties of bovine oligosaccharide-binding protein. A microbicidal granule protein of		

JOURNAL	eosinophils and neutrophils					
MEDLINE	J. Biol. Chem. 277 (22), 19658-19664 (2002)					
PUBMED	22028028					
REFERENCE	11860375					
AUTHORS	2 (bases 1 to 688)					
TITLE	Yount,N.Y., Yuan,J., Tydeall,C.C. and Selsted,M.E.					
JOURNAL	Direct Submission					
FEATURES	Submitted (11-MAR-2002) Pathology, UC Irvine, 440D Medical Sciences 1, Irvine, CA 92697, USA					
SOURCE	Location/Qualifiers 1..688					
CDS	/organism="Bos taurus" /db_xref="taxon:9913" /cell_type="peripheral white blood cells" 29..601 /note="microbicidal; undergoes N-terminal glutamine cyclization; peptidoglycan recognition protein; pgrp"					
misc_feature	/product="Oligosaccharide-binding protein" /codon_start=1 /protein_id="BAL87002.1" /db_xref="GI:19550242"					
misc_feature	/translation="MSRRPTPLAVALLIGLGAADCGSIYSRCKKGLALSKCSORL ROPVIVVSHRAGSYCTPASPASCORANOVITYHVRGWCVDVYNLLIEDGLVYG KGWNTGASGPTNRPFLAIGISFGMTWHRYAPRAAQSILAAGARGYLTPNEY VKGHADVQDTSPGGELTKIIIOGPHRV"					
misc_feature	order(98..100,470..472) /note="disulfide bond"					
misc_feature	order(146..148,281..283) /note="disulfide bond"					
misc_feature	order(209..211,227..229) /note="disulfide bond"					
BASE COUNT	126 a	234 c	201 g	127 t		
ORIGIN						
Alignment Scores:						
Pred. No.:	7,82e-61	Length:	688			
Score:	717.00	Matches:	131			
Percent Similarity:	80.75%	Conservative:	20			
Best Local Similarity:	70.05%	Mismatches:	30			
Query Match:	68.22%	Indels:	6			
DB:	4	Gaps:	1			
US-09-462-625-4 (1-191) x AY083309 (1-688)						
Oy	3	LeuaLATPALaleProSerleuLeuaRgyLeugLYalaAlaIngLUThrGlaspPro	22			
Dd	50	CTCGCCTGGGTCCTCTGTGCCCTCTCTGGCCCTCGGGCGGCTCAAGC-----	97			
Oy	23	AlAcysCySerProIleValPrOAryasnLuTrrpySaLaLeuaAlasergLucyAla	42			
Dd	98	-----TGCgcAGCATGTGTccCGcGGAAGTggGgCcCCTGGcATCCAAgtCGACG	151			
Oy	43	GINHisLeuSerLeuProLeuaRgyTyVaLYalValSerHisThrAlagLySerSerCys	62			
Dd	152	CAGAGCGCTAACACAACCCTGTGCTAGCTGGTGTTGTCGCACAGCGGGCACCGCTGC	211			
Oy	63	AsnThrProAlasercYsgLINGlInglAlalargsanYalGINHisTryHisMetLYthr	82			
Dd	212	AACACTCCGGGCTGTGGTCAGAGCAAGGCCAAAACGTGCAGTACTACCAcGTGcgGAG	271			
Oy	83	LeugLTrrPCysAspValGlyTYTrAsnPhelEuILecylglunspglyLeuValTYrglu	102			
Dd	272	CGGGCTGTGTCCGACGTGGGTCTACAAATTCTCTGATCGAGAGAATGGGGCTGTGTAG	331			
Oy	103	GLYAtrgLYTrPanPhetThrGLYAlAHIserserLYHisLeutrPanPrrMetserIle	122			
Dd	332	GGCGGGGGCTGGAAcACCTTAGGTGTCTACTGTGGGCCCAcGTGGAACCCCAATAGCCATC	391			
Oy	123	GLYIlleserPhemetGLyasnTYrMetAspArgVALProThrProGINAlaiIleatgALA	142			
Dd	392	GGCATTCCTTCATGGGCAACTACATGCANCGGGGTGCCCGCGGCTGTGCTCAGGGGS	451			
Oy	143	AlaInglnLYleuLeuaIacySGlyValAlaInglnLYlaLeuArSeraSnTYrValLeu	162			

[illegible]

DB	181	TTCGTGAAACAGCCCGGACTCTCTGTGAACAGCAGGCCCGCAATGTGCAGCATTCACCAAG	240
OY	81	LysThrLeuLyrPCysAspValGLYTyrAsnPhenuleuIleuLyGuaspGLyLeuVal	100
DB	241	AARGACTGGGCGTGTGGCGATTGACCTTCAACACTTCTTAATGGAGAAGCGGTCATGTC	300
OY	101	TyrGLuGLyARgSLYTrpAsnPherHrGLyAlAHISerGLYHisLeuTrpAsnPromet	120
DB	301	TATGAAGGCCCGAGCTGTGAACATCAACAGGCCGACCACAGAGGCCCATCTCGAATCCCANG	360
OY	121	SerILeGLYLLeSerPheMetGLYasnTYrMetasParValProThrProGlnalaile	140
DB	361	TCATTGGCATCACCTTCATGGGAACTTCATGAGCCGGGTACCCCAAGGGGCCCTC	420
OY	141	ArgAlaAlaGLInGLyLeuLeuAlaCySGLyValAlaIGLnGLyAlaLeuArgerSerAsnTYr	160
DB	421	CGTGCGCCCTAAATCTTGTGGAATGTGGGGGTCTCTCGGGGCTTCTCGATCCAATAT	480
OY	161	ValLeuLeuGLYGLYHisARgAspVALGlnArghrLeuSerProGLYAsnGlnLeuTyrlHis	180
DB	481	GAGGTCAAAGGACACCGGATGTGCAAGACACTCTCTCTCCAGHGACCAACTCATATCAG	540
OY	181	LeuILeGlnAsnTProHistrTYrarng	189
DB	541	GTCAATCAAGCTGGGACACTACCGA	567
RESULT 7			
AF076482			
LOCUS	AF076482	680 bp	mRNA linear ROD 15-AUG-1998
DEFINITION	Mus musculus peptidoglycan recognition protein precursor (Pgpr)		
ACCESSION	AF076482		
VERSION	AF076482.1		
KEYWORDS	GI:3342530		
SOURCE	Mus musculus.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H. A peptidoglycan recognition protein in innate immunity conserved from insects to humans Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998) 98374308 9707603		
JOURNAL	2 (bases 1 to 680)		
MEDLINE	Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H. Direct Submission Submitted (08-JUL-1998) Microbiology, Stockholm University, Stockholm S-106 91, Sweden		
REFERENCE	Location/Qualifiers		
AUTHORS	1. 680		
JOURNAL	/organism="Mus musculus"		
TITLE	/db_xref="taxon:10090"		
PUBMED	/tissue_type="spleen"		
FEATURES	1. 680		
source	/gene="Pgpr"		
	36..584		
gene	/gene="Pgpr"		
CDS	/note="Pgpr"		
	/codon_start=1		
	/product="peptidoglycan recognition protein precursor"		
	/protein_id="AAC31821.1"		
	/db_xref="GI:3342531"		
	/translation="MLFAALAILGLASCSFIYPRSRRLPSQSSRLGHPIRYV VISHTSGRSCNSPDSCEDQARNVQHTHKNELCMQDVAYNFLIGEDGHVEGRWNITKG DHTRGFPMNWSIGITFNGFMNDRPAPAKRALRAALMLECVSRGFLRSNYEYVGHRDY QSLSFGDOLYGVIOSWEHYRE"		
	36..83		
sig_peptide	/gene="Pgpr"		
	84..581		
mat_peptide	/gene="Pgpr"		
	/product="peptidoglycan recognition protein"		

QY 101 TyGluGlyArgGlyTyrPaspnPhetHrglyAlaHisSerGlyHisLeuTyrPaspnProket 120
 Db 323 TATGAAGGCGAGGCTGGAAATCATCAAGGTCACACAGGCGCCCATGTGAAATCCCATG 382
 QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 Db 383 TCTATTGGCACTACCTTCATGTGGAGACCTTCACCGGATACCGCAAGCGGGCCCTC 442
 QY 141 ArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
 Db 443 CGTCTGCGCCCAATATCTTTCGAATGTGGGTGTCTGGGCTTCTTGAAATCCAACTAT 502
 QY 161 ValLeuGlySerGlyHisArgAspValAlaArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 Db 503 GAAGTCAAGAGACACCGGAGATGTGCAAGACACTCTCTCTCCAGGTGACCACTATACG 562
 QY 181 LeuIleGlnAsnTyrProHisTyrArg 189
 Db 563 GTCAATCCAAAGCTGGGAACACTACCGA 589
 RESULT 9
 AF154114 630 bp mRNA linear ROD 01-JUN-2000
 LOCUS Rattus norvegicus peptidoglycan recognition protein PGRP (Pgrp)
 DEFINITION mRNA, complete cds.
 ACCESSION AF154114
 VERSION AF154114.1 GI:8132325
 KEYWORDS
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 630)
 Rehman, A., Teedecki, E. E. and Krueger, J. M.
 Direct Submission
 Submitted (25-MAY-1999) Veterinary Comparative Anatomy,
 Pharmacology, and Physiology, Washington State University, P.O. Box
 646520, Pullman, WA 99164, USA
 Location/Qualifiers
 1..630
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /tissue_type="spleen"
 1..630
 /gene="Pgrp"
 20..571
 /gene="Pgrp"
 /note="similar to Homo sapiens and Mus musculus
 peptidoglycan recognition protein; similar to Trichoplusia
 ni and Mus musculus TAG7"
 /codon_start=1
 /product="peptidoglycan recognition protein PGRP"
 /protein_id="AAFP3252.1"
 /db_xref="GI:8132326"
 /translation="MLFAWAPPALGLADSCFYVRESEKALPSECKGLKPKYR
 VVISHAGFCSSPDSCQEQARNVOLYQKQKQDVAAYNFIIGEDGHYERGMATIK
 GDHTGPIWNPMSIGITFMGDYSHRYPARAKALAAALNLKCGVSEGLRNVYKRGHRD
 VOSTLSPDQLEILOSMDHYRE"
 BASE COUNT 134 a 193 c 166 g 137 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.7e-56 Length: 630
 Score: 667.50 Matches: 124
 Percent Similarity: 77.37% Conservative: 23
 Best Local Similarity: 65.26% Mismatches: 34
 Query Match: 63.51% Indels: 9
 DB: 10 Gaps: 3
 US-09-462-625-4 (1-191) x AF154114 (1-630)

QY 1 MetLeuLeuAlaTyrPala---LeuProSerLeuLeuArgLeuGlyAlaAlaGlnGlyThr 19
 Db 20 ATGTGTGTTGCTGGGCTGCTCCCTGCTGCTCCCTGGGCGCTGGCA----- 64
 QY 20 GlnAspProAlaCysCysSerProIleValProArgAsnGlyTyrPalaLeuAlaSer 39
 Db 65 -----GACCTCTGCTGTTTC--GTCGTGCCCCCAGCTGAGTGGAGGCCCTGCCATTC 115
 QY 40 GluCysAlaGlnHisLeuSerLeuProLeuArgTyrValAlaValSerHisThrIleGly 59
 Db 116 GAGTGTCTCAAGGGCGCTGAAGAAGCACTCCGCTACGCTGGTGAATTCACACACAGCCGCG 175
 QY 60 SerSerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisTyrHis 79
 Db 176 AGCTTCTGCACACACCCAGACTCTCTGTGACAGAGGCCCAATGTGCACCTTACCAA 235
 QY 80 MetTyrThrLeuGlyTyrCysAspValAlaGlyTyrAsnPhetHrglyAlaHisSerGlyHisLeuTyrPaspnPro 99
 Db 236 ATGAAGACGCTGGGCTGGTGGATGTACCTTACCACTTCATTTGGAGAGATGTCTAT 295
 QY 100 ValTyrGluGlyArgGlyTyrPaspnPhetHrglyAlaHisSerGlyHisLeuTyrPaspnPro 119
 Db 296 GTTACGAGAGCGGAGGCTGGACCATCAAGGTGACACACAGGCGCCATGTGAAACCC 355
 QY 120 MetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAla 139
 Db 356 ATGCTTACGCGACACCTTCATGTGATGACACTACACCGGATACCTGCAAGCGGGCT 415
 QY 140 IleArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsn 159
 Db 416 CTCGCTGCTGCGCCCAATATCTTCTGAATGTGGGTGTCTGGGCTTCTTGAGATCTTAC 475
 QY 160 TyrValLeuGlySerGlyHisArgAspValAlaArgThrLeuSerProGlyAsnGlnLeuTyr 179
 Db 476 TATGAGGTCAAGAGACATCAGATGTGCAAGACACTCTGTCTCCAGGTGACCACTTAC 535
 QY 180 HisLeuIleGlnAsnTyrProHisTyrArg 189
 Db 536 GAGATCAATCCAAAGCTGGGAACACTACCGA 565
 RESULT 10
 MMRNATMS1 678 bp mRNA linear ROD 30-JUL-1998
 LOCUS M.musculus mRNA for TAG7 protein.
 DEFINITION X86374
 ACCESSION X86374.1 GI:887524
 VERSION X86374.1
 KEYWORDS tms1 gene.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 1 (bases 1 to 678)
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 Afanas'eva, A.V. and Kabishhev, A.A.
 Cloning of the tag7 gene expressed in metastatic mouse tumors
 Genetika 32 (5), 621-628 (1996)
 8753036
 2 (bases 1 to 678)
 Kabishhev, A.A., Iukandidin, E.M. and Georgiev, G.P.
 Direct Submission
 Submitted (18-APR-1995) O.S. Kustikova, Inst of Gene Biology
 Russian Academy, of Science, B334, Vavilov Street 34/5, Moscow,
 RUSSIA
 3 (bases 1 to 678)
 Kiselev, S.L., Kustikova, O.S., Korobko, E.V., Prokhortchouk, E.B.,
 Kabishhev, A.A., Iukandidin, E.M. and Georgiev, G.P.
 Molecular cloning and characterization of the mouse tag7 gene
 encoding a novel cytokine
 J. Biol. Chem. 273 (29), 18633-18639 (1998)
 JOURNAL MEDLINE 98325081
 PUBMED 9660837
 COMMENT On Jul 6, 1995 this sequence version replaced gi:785010.


```

misc_feature
      HOADMATIDARLALQRYMNRDMS"
      746..904
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (12..170): 99% identity.-AA283147 z149d07.r1 NCI_CGAP_GCB1
      Homo sapiens cDNA clone IMAGE:713101 5'; (6..163); 99%
      identity.-AA143160 z149d07.r1 Soares pregnant uterus NBHPU
      Homo sapiens cDNA clone 505261 5'; (1..43); 100%
      identity.-AA142881 z149d07.s1 Soares pregnant uterus NBHPU
      Homo sapiens cDNA clone 505261 3'; (570..527); 64%
      identity."
      1157..1239
      /note="Predicted exon, program: gra112exons_human_1.3,
      frame: 0, quality: good, score: 62.000-DSD similarity to
      overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (171..255): 95% identity.-AA283147 z149d07.r1
      NCI_CGAP_GCB1 Homo sapiens cDNA
      clone IMAGE:713101 5'; (164..247); 92% identity.-AA143160
      z149d07.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA
      clone 505261 5'; (44..126); 100% identity.-N28707
      yx66d11.r1 Homo sapiens cDNA clone 266709 5'; (1..66); 98%
      identity.-AA142881 z149d07.s1 Soares pregnant uterus NBHPU
      Homo sapiens cDNA clone 505261 3'; (526..438); 82%
      identity."
      1536..1634
      /rpt_family="(CA)n"
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (1673..1973)
      /rpt_family="AluX"
      complement(2040..2185)
      repeat_region
      /rpt_family="LINE2"
      complement(2200..2498)
      repeat_region
      /rpt_family="AluX"
      complement(2576..2746)
      repeat_region
      /rpt_family="AluSg/x"
      complement(2747..3044)
      repeat_region
      /rpt_family="AluX"
      3048..3750
      repeat_region
      /rpt_family="MER64"
      3841..4060
      repeat_region
      /rpt_family="MER64"
      4294..4451
      misc_feature
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (256..438); 81% identity.-AA283147 z149d07.r1
      NCI_CGAP_GCB1 Homo sapiens cDNA
      clone IMAGE:713101 5'; (248..400); 99% identity.-AA143160
      z149d07.r1 Soares pregnant uterus NBHPU Homo sapiens cDNA
      clone 505261 5'; (127..281); 97% identity.-N28707
      yx66d11.r1 Homo sapiens cDNA clone 266709 5'; (67..222);
      98% identity.-AA142881 z149d07.s1 Soares pregnant uterus
      NBHPU Homo sapiens cDNA clone 505261 3'; (437..275); 97%
      identity."
      complement(4789..5083)
      repeat_region
      /rpt_family="AluX"
      5875..5972
      misc_feature
      /note="DSD similarity to overlapping ESTs:
      N39722 yx92e04.r1 Homo sapiens cDNA clone 269214 5';
      (439..506); 71% identity.-AA143160 z149d07.r1 Soares
      pregnant uterus NBHPU Homo sapiens cDNA clone 505261 5';
      (282..379); 100% identity.-AA142881 z149d07.s1 Soares
      pregnant uterus NBHPU Homo sapiens cDNA clone 505261 3';
      (274..175); 95% identity."
      5875..5898
      misc_feature
      /note="DSD similarity to N28707 yx66d11.r1 Homo sapiens
      cDNA clone 266709 5'; (223..246); 100% identity."
      5906..6036
      misc_feature
      /note="Predicted exon, program: gra112exons_human_1.3,
      frame: 1, quality: excellent, score: 100.000"
      complement(6148..6443)
      repeat_region
      /rpt_family="AluX"
      complement(6608..6648)
      repeat_region
      /rpt_family="LINE2"

repeat_region
      complement(7118..7340)
      /rpt_family="MIR"
      repeat_region
      complement(7538..7639)
      /rpt_family="(CA)n"
      repeat_region
      complement(7792..8054)
      /rpt_family="MIR"
      repeat_region
      complement(8789..9083)
      /rpt_family="AluJo"
      repeat_region
      9456..9520
      /rpt_family="MER5B"
      repeat_region
      complement(9539..9841)
      /rpt_family="AluX"
      repeat_region
      9909..9944
      /rpt_family="MER5B"
      repeat_region
      10099..10401
      /rpt_family="AluX"
      repeat_region
      complement(10587..10752)
      /rpt_family="AluJ"
      repeat_region
      complement(10753..11057)
      /rpt_family="AluX"
      repeat_region
      complement(11065..11191)
      /rpt_family="AluJ"
      repeat_region
      11322..11619
      /rpt_family="AluSg"
      repeat_region
      11775..11814
      /rpt_family="Ar_r1ch"
      repeat_region
      complement(11864..11885)
      /rpt_family="Ar_r1ch"
      repeat_region
      complement(11914..11976)
      /rpt_family="AluS"
      repeat_region
      complement(12055..12088)
      /rpt_family="Ar_r1ch"
      repeat_region
      complement(12186..12327)
      /rpt_family="LINE1"
      repeat_region
      complement(12334..12384)
      /rpt_family="LINE2"
      12784..12994
      misc_feature
      /note="Predicted exon, program: gra112exons_human_1.3,
      frame: 1, quality: excellent, score: 92.000"
      12847..12908
      repeat_region
      /rpt_family="GC_r1ch"
      12895..12994
      misc_feature
      /note="DSD similarity to AA765166 n279c05.s1 NCI_CGAP_GCB1
      Homo sapiens cDNA clone IMAGE:1301672; (222..123); 97%
      identity."
      13002..13032
      repeat_region
      /rpt_family="GC_r1ch"
      13080..13162
      misc_feature
      /note="Predicted exon, program: gra112exons_human_1.3,
      frame: 2, quality: excellent, score: 100.000-DSD
      similarity to AA765166 n279c05.s1 NCI_CGAP_GCB1 Homo
      sapiens cDNA clone IMAGE:1301672; (122..48); 100%
      identity."
      complement(13799..13954)
      repeat_region
      /rpt_family="GC_r1ch"

Alignment Scores:
Pred. No.: 1.45e-37 Length: 166500
Score: 505.50 Matches: 100
Percent Similarity: 73.72% Conservative: 1
Best Local Similarity: 72.99% Mismatches: 0
Query Match: 48.10% Indels: 36
DB: 9 Gaps: 1

US-09-462-625-4 (1-191) x AC007785 (1-166500)
QY 91 AsnPhenLeuIleGlyIAspGlyLeuValTyroGlyArgGlyTPasPhenrGly 110
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 17384 AGCTTCTGATTGTGAGAAAGCGGCTGCTATPACAGAGGCGTGCTGGAATTCACGGCT 17325

QY 111 AlAhHisSerGlyHisLeuTPasnPrometSerIleGlyIleSerPhenMetGlyAsnTy 130
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 17324 GCCCAGTCAGAGTCATTATGGAACCCCATCTCCATGTGCATACCTTCATGGCAACTAC 17265

```


QY 131 Met----- 131
 Db 17264 ATGGGTAGTAGTGTCCAGCAGTGGATGGGGGTGACAGGGGCCCTACGATCA 17205
 QY 132 -----AsparValPr 135
 Db 17204 CTTCCTGTGGGGAGGCCCTGCCTGACCTCTACCTCTGCTCTCCAGATCGGGTGC 17145
 QY 135 CThPrProGlnAlaIleArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAl 155
 Db 17144 CACACCCAGGCGCATCCGGGAGGCCAGGCTCTACTGCTGCGGTGTGGCTCAGGAGC 17085
 QY 155 AlenArgSerAntyValLeuLysGlyHisArgAspValGlnArgThrLeuSerProG 175
 Db 17084 CTTAGGTCCACTAGTATGTCTCAAGGACCGGATGTGACGCTACACTCTCCAGG 17025
 QY 175 YasnGlnLeuTyrlHisLeuIleGlnAsnTrpProHisTyrArgSerPro 191
 Db 17024 CAACGAGCTGTACACCTCATCCAGAAATTGGCCACACTACCGCTCCCC 16976
 RESULT 12
 LOCUS AY035376
 DEFINITION Homo sapiens peptidoglycan recognition protein-I-alpha precursor
 ACCESSION AY035376
 VERSION AY035376
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1173)
 AUTHORS Liu, C., Xu, Z., Gupta, D. and Dziarski, R.
 TITLE Peptidoglycan recognition proteins: a novel family of four human
 JOURNAL Innate immunity pattern recognition molecules
 MEDLINE J. Biol. Chem. 276 (37), 34686-34694 (2001)
 PUBMED 21433985
 11461926
 2 (bases 1 to 1173)
 REFERENCE Liu, C., Xu, Z., Gupta, D. and Dziarski, R.
 AUTHORS Direct Submission
 TITLE Submitted (14-MAY-2001) Northwest Center for Medical Education,
 JOURNAL Indiana University School of Medicine, 3400 Broadway, Gary, IN
 46408, USA
 FEATURES
 source Location/Qualifiers
 1. 1173
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q21; BAC 236c22"
 1. 1173
 /gene="PGLYRP1alpha"
 /note="PGRP1alpha"
 54. 1079
 /gene="PGLYRP1alpha"
 /note="PGRP-1 alpha; member of PGRP family; innate
 immunity pattern recognition molecule; contains 2
 transmembrane domains, 2 extracellular domains, and 1
 cytoplasmic domain"
 /codon_start=1
 /product="peptidoglycan recognition protein-I-alpha
 precursor"
 /protein_id="AAK72484.1"
 /db_xref="GI:15590684"
 /translation="MGILPMLAFPIIGLQAMDPTIYSRKEMGARPLACRALTLIPV
 AYITIDLPGMOCOSVCSOMLRFGLIOASHVYITGMDVAVNPLVGDGRYRGVGN
 IGLHTDGYNNISLIGLAFPGKRTSSSPALSAEGLISITAIKSHSPRTIPLLL
 KEETCLDHPHVRKVCNPNTIKRSAMEARETHCPKNNLPARKYIILHTAGTSCYST
 KCOTVVRNIOFSHMDTNCFDIGYHFLVGDGCVYEGVMHIGSHRYGFDNLGIA
 FIGYFVERPNAALFAAODLIGCAVVEGYLPTNVLMLGSHDVNIIISPQALNIIIS
 TWPFRFK"

sig_peptide 54. 104
 /gene="PGLYRP1alpha"
 misc_feature 294. 377
 /gene="PGLYRP1alpha"
 /note="Region: conserved PGRP domain IV"
 misc_feature 642. 686
 /gene="PGLYRP1alpha"
 /note="Region: conserved PGRP domain III"
 misc_feature 762. 848
 /gene="PGLYRP1alpha"
 /note="Region: conserved PGRP domain II"
 misc_feature 921. 1070
 /gene="PGLYRP1alpha"
 /note="Region: conserved PGRP domain I"
 BASE COUNT 250 a 364 c 297 g 262 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,08e-33 Length: 1173
 Score: 441.00 Matches: 77
 Percent Similarity: 64.52% Conservative: 43
 Best Local Similarity: 41.40% Mismatches: 60
 Query Match: 41.96% Indels: 6
 DB: 9 Gaps: 3
 US-09-462-625-4 (1-191) x AY035376 (1-1173)
 QY 8 ProSerLeuLeuArgLeuGlyAlaIleGlnGlnThrGlnAspProAla----- 23
 Db 522 CCACCTCTCTGAAGAAGAGACGCTGCTGACCCCTCAACATCCAGTATGCCAGGAG 581
 QY 24 CysCysSerProIleValProArgAsnGlnTyrPlyAlaLeuAlaSerGlyCysAlaGln 43
 Db 562 GTTTCCTCCCAATCATCAACGATCTGCTGGAGCCAGAGACACACTGCCCTAA 641
 QY 44 HisLeuSerLeuProLeuArgTyrValAlaValAlaSerHisThrAlaGlySerCysAsn 63
 Db 642 ---ATGACCTCCAGCCCAAAATGTGCATCATCATCCACACCGCTGCCACACCTCACT 698
 QY 64 ThrProAlaSerCysGlnGlnGlnAlaIleArgAsnValGlnHisTyrHisMetLysThrLeu 83
 Db 699 GTATCCACAGACTGCACGACTGCTGCGCAAAATGACATGCTCTTGCATGACACAGG 758
 QY 84 GlyTrpCysAspValGlyTyrAsnPhleuIleGlyGlnAspGlyLeuValTyrGlnGly 103
 Db 759 AACTTTGTGACATGAGATCATCTTCTGTGGGCGAGAGTGGCGGTATGAAGG 818
 QY 104 ArgGlyTrpAsnPhleuThrGlyAlaIleHisSerGlyHisLeuTrpAsnProMetSerIleGly 123
 Db 819 GTTGATGGCACAATCCAGAGCTCTCACT---TAGGATTCACAGATATATGCCCTAGGA 875
 QY 124 IleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIleArgAlaAla 143
 Db 876 ATTGCTTCATCGCTACTTGTGTAGAAAGCTCCAAATGTGTCACAGCTGGAGCGCGCC 925
 QY 144 GlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuAlaArgSerAsnTyrValLeuLys 163
 Db 936 CAGGACCTGATCCAGTGCCTGCTGTGGAGGGGTACTCACTCCAAACATCACTGCGATG 995
 QY 164 GlyHisArgAspValGlnArgThrLeuSerProGlyAlaGlnGlyLeuTyrlHisLeuIleGln 183
 Db 996 GCCCAGAGAGAGCTGTCAACATCTGTCTCCCTGGGAGAGCTTTGTATTAACATCATCAGC 1055
 QY 184 AsnTrpProHisTyrArg 189
 Db 1056 ACCTGGCCCATTTCAAG 1073
 RESULT 13
 LOCUS AC007085
 DEFINITION AC007085
 ACCESSION AC007085
 166867 bp DNA linear INV 17-MAR-2001
 Drosophila melanogaster, chromosome 2R, region 44D2-44E1, BAC clone
 BACR21H10, complete sequence.

[illegible]

Wed Nov 13 09:37:09 2002

us-09-462-625-4.rge

Page 13

```
QY      188  TyrArgSer  190
          :::::::::::
Db 150729 TGGAGGCC 150737
```

Search completed: November 12, 2002, 05:02:37
Job time : 2802 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 00:51:04 : Search time 85 Seconds

(without alignments)
93.200 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAMALPSLRGAAQETE.....LSPGNQLYHLQNPVHRSP 191

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1051	100.0	196	1	PGRP_HUMAN
2	692	65.8	182	1	PGRP_MOUSE
3	388.5	37.0	182	1	PGRP_TRINI
4	345.5	32.9	196	1	PGRP_BOMMO
5	127	12.1	150	1	NAAA_BPT7
6	127	12.1	150	1	NAAA_BPT7
7	86.5	8.2	481	1	ATZB_PSED
8	79.5	7.6	267	1	SURE_HELPJ
9	79.5	7.6	267	1	SURE_HELPJ
10	77.5	7.4	356	1	YGTJ_ECOLI
11	75.5	7.2	443	1	FBIA_CRIGR
12	75	7.1	687	1	VS41_GIALA
13	75	7.1	764	1	GLGB_STRAU
14	75	7.1	803	1	MDDC_CHLIR
15	74	7.0	713	1	TS44_GIALA
16	74	7.0	1477	1	HTK7_HYDAT
17	73.5	7.0	439	1	YSCN_YEREN
18	73.5	7.0	439	1	YSCN_YEREN
19	73.5	7.0	1733	1	VNUA_PRYCA
20	72.5	6.9	378	1	CARA_PSEAE
21	72	6.9	207	1	SNAT_SHEEP
22	72	6.9	408	1	G58B_DROME
23	72	6.9	1286	1	IRBP_BOVIN
24	71.5	6.8	362	1	IRBP_BOVIN
25	71.5	6.8	444	1	IRBP_BOVIN
26	71	6.8	402	1	GLGA_DEIRA
27	71	6.8	775	1	LIPS_HUMAN
28	71	6.8	1120	1	YB85_YEAST
29	71	6.8	1123	1	DC13_DROME
30	71	6.8	1589	1	DC13_DROME
31	70.5	6.7	443	1	FBIA_MOUSE
32	70.5	6.7	610	1	GIDA_CHLTR
33	70.5	6.7	3312	1	CLR3_HUMAN

34	70	6.7	402	1	PGT1_ERMCA	P18192 erwinia car
35	70	6.7	644	1	YL53_HUMAN	O14166 homo sapien
36	70	6.7	883	1	PGCB_MOUSE	O61361 mus musculu
37	70	6.7	1216	1	AEGB_RAT	O63191 ratu mus norv
38	69.5	6.6	434	1	PTN1_CHICK	O13016 gallus gall
39	69.5	6.6	700	1	PTPE_HUMAN	P23469 homo sapien
40	69.5	6.6	818	1	PTK2_YEAST	P47116 saccharomyc
41	69	6.6	676	1	NTG6_RICCN	O92692 rickettsia
42	69	6.6	726	1	BP1L_HUMAN	P50747 h biotin--p
43	69	6.6	759	1	LIPS_MOUSE	P54310 mus musculu
44	69	6.6	1033	1	CR2_HUMAN	P20023 homo sapien
45	69	6.6	1436	1	WC11_BOVIN	P30205 bos taurus

ALIGNMENTS

RESULT 1
ID PGRP_HUMAN STANDARD: PRT; 196 AA.
AC 075594; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptidoglycan recognition protein precursor (SBB168).
GN PGLYRP OR PGRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE-Bone marrow;
RX MEDLINE=98374508; PubMed=9707603;
RA Kang D., Liu G., Lundstroem A., Gellius E., Steiner H.;
RT "A peptidoglycan recognition protein in innate immunity conserved from
insects to humans.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998). *Ang 18*
RN [3]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Richardson P., Sakaladasis G.,
RA Burkhart-Schultz K., Gordon L., Scott D., Johnson G., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Dangnan L.,
RA Erlar A., Christensen M., Georgescu A., Avila J., Attix C.,
RA Andreise T., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Ariellano A., Sanders C.,
RA Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.,
RN Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Man T., Zhang W., Gao X.;
RA Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RL - FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN
INNATE IMMUNITY.
CC - FUNCTION: THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO (BY
SIMILARITY).
CC - SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-
ASSOCIATED FORMS (BY SIMILARITY).
CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BONE MARROW, WEAK
EXPRESSION FOUND IN KIDNEY, LIVER, SMALL INTESTINE, SPLEEN,
THYMUS, PERIPHERAL LEUKOCYTE, LUNG AND FETAL SPLEEN.
CC - SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
CC EMBL; AF076483; AAC31822.1; -

DR EMBL: AC007785; AAD38243.1; -
 DR EMBL: AF242517; AAF95958.1; -
 DR HSSP: P00806; IABO.
 DR Genew: HGNC:8904; PGLYRP.
 DR MIM: 604963; -
 KW Immune response; Cytokine; Apoptosis; Signal.
 FT SIGNAL 1 21
 FT CHAIN 196 196
 FT DISULFID 67 73
 FT CARBOHYD 112 112
 FT SEQUENCE 196 AA; 21731 MW; D954C51440DC27DC CRC64;
 SQ
 Query Match 100.0%; Score 1051; DB 1; Length 196;
 Best Local Similarity 100.0%; Pred. No. 5; 3e-94;
 Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLALALPSILRLGAAOETEDPACCSPTVPRNEMKALASECAQHLSPRLRYVVSHTAGS 60
 DB 6 MLALALPSILRLGAAOETEDPACCSPTVPRNEMKALASECAQHLSPRLRYVVSHTAGS 65
 QY 61 SCNTPASCOOQARVVOHYHMKTLGMCVDVGNFLIGEDGLYEGRGWNTGAHSGHLNMPM 120
 DB 66 SCNTPASCOOQARVVOHYHMKTLGMCVDVGNFLIGEDGLYEGRGWNTGAHSGHLNMPM 125
 QY 121 SIGISFMGNMTDRVPTPOAIRAAQGLACGVAQGLRSNYVLKGRDVOFTLSFGNOLYH 180
 DB 126 SIGISFMGNMTDRVPTPOAIRAAQGLACGVAQGLRSNYVLKGRDVOFTLSFGNOLYH 185
 QY 181 LIONMHRPSP 191
 DB 186 LIONMHRPSP 196
 RESULT 2
 PGRP MOUSE STANDARD; PRT; 182 AA.
 ID PGRP MOUSE STANDARD; PRT; 182 AA.
 AC 088393; 062185;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Peptidoglycan recognition protein precursor (Cytokine tag7).
 GN PGLYRP OR PGRP OR TAG7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_Taxid-10090;
 OX [1]
 RN SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RP TISSUE-Spleen;
 RC MEDLINE-98374308; PubMed-9707603;
 RX Kang D., Liu G., Lundstrom A., Gellius E., Steiner H.;
 RT "A peptidoglycan recognition protein in innate immunity conserved from
 insects to humans";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).
 RN [2]
 RN SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RP MEDLINE-98325081; PubMed-9660837;
 RX Kiselev S.L., Kustikova O.S., Korobko E.V., Prokhorchouk E.B.,
 RA Kabishhev A.A., Lukandin E.M., Georgiev G.P.;
 RT "Molecular cloning and characterization of the mouse tag7 gene
 encoding a novel cytokine";
 RL J. Biol. Chem. 273:18633-18639(1998). *July 17*
 RN [3]
 RN SEQUENCE FROM N.A.
 RP Slayton W.B., Rigas A., Hancock J.D., Zaugg J.K., Le T.V.,
 RA Trantman M.S., Spangrude G.J., Carroll W.L., Schibler K.R.;
 RT "Granulocyte-colony stimulating factor up-regulates expression of
 murine tag7 during myeloid differentiation";
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN-C57BL/6J; TISSUE-Small intestine;
 RX MEDLINE-21085660; PubMed-11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa T., Hara A., Fukunishi Y., Konno H., Aachji J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann M., Gaasterland T., Gissi C., King B., Kochia H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombers P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilmink L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP Strauberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN
 CC INNATE IMMUNITY. THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO.
 CC -1- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-
 CC ASSOCIATED FORMS.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN AND LUNG. ALSO
 CC DETECTED IN BRAIN AND THYMUS. IN THE LUNG, EXPRESSED IN THE
 CC INTRAALVEOLAR SPACE. IN THE BRAIN, EXPRESSED IN THE PURKINJE CELLS
 CC OF THE CEREBELLUM AND IN CERTAIN LAYERS OF NEURONS IN THE
 CC HIPPOCAMPUS. ALSO DETECTED IN CELLS FILLING THE SPACE WITHIN THE
 CC INTESTINAL VILLOS.
 CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
 CC FAMILY.
 CC [6]
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: AF076482; AAC31821.1; -
 DR EMBL: AF193843; AAF06335.1; -
 DR EMBL: AK008335; BAB25611.1; -
 DR EMBL: BC005582; AAH05582.1; -
 DR EMBL: X86374; CAAG0133.1; ALT_SEQ.
 DR EMBL: Y12088; CAA72803.1; -
 DR MGD; MGI:1345092; Pglyrp.
 KW Immune response; Cytokine; Apoptosis; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 182
 FT DISULFID 54 60
 FT SEQUENCE 182 AA; 20489 MW; 9844E2137F047F14 CRC64;
 SQ
 Query Match 65.8%; Score 692; DB 1; Length 182;
 Best Local Similarity 67.2%; Pred. No. 1; 6e-59;
 Matches 127; Conservative 21; Mismatches 33; Indels 8; Gaps 1;
 QY 1 MLALALPSILRLGAAOETEDPACCSPTVPRNEMKALASECAQHLSPRLRYVVSHTAGS 60
 DB 1 MLALALPSILRLGAAOETEDPACCSPTVPRNEMKALASECAQHLSPRLRYVVSHTAGS 52
 QY 61 SCNTPASCOOQARVVOHYHMKTLGMCVDVGNFLIGEDGLYEGRGWNTGAHSGHLNMPM 120
 DB 53 SCNTPASCOOQARVVOHYHMKTLGMCVDVGNFLIGEDGLYEGRGWNTGAHSGHLNMPM 112
 QY 121 SIGISFMGNMTDRVPTPOAIRAAQGLACGVAQGLRSNYVLKGRDVOFTLSFGNOLYH 180
 DB 113 SIGISFMGNMTDRVPTPOAIRAAQGLACGVAQGLRSNYVLKGRDVOFTLSFGNOLYH 172

OY 181 LIONMPHYR 189
DB 173 VIOSMEHYR 181

RESULT 3

PGRP_TRINI STANDARD; PRT; 182 AA.

AC 075537; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptidoglycan recognition protein precursor.

OS Trichoplusia ni (Cabbage looper).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Noctuoidea; Noctuidae; Plusiinae; Trichoplusia.
OX NCBI_Taxid=7111;

RP SEQUENCE FROM N.A., SEQUENCE OF 17-21, TISSUE SPECIFICITY, AND
INDUCTION.

RC MEDLINE-98374308; PubMed-9707603;
RA Kang D., Liu G., Lundstroem A., Gellius E., Steiner H.;
RT "A peptidoglycan recognition protein in innate immunity conserved from
insects to humans."
RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).

CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND TRIGGERS THE
PROPODOLIXIDASE CASCADE WHICH IS AN IMPORTANT INSECT INNATE IMMUNE
DEFENSE MECHANISM.

CC -1- SUBUNIT: MONOMER (PROBABLE).
CC EXPRESSION OBSERVED IN HEMOCYTE. NO EXPRESSION DETECTED IN
GUT.

CC -1- INDUCTION: BY BACTERIAL CHALLENGE.
CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF076481; AAC31820.1; -
KW Immune response; Signal.

FT SIGNAL 1 16
FT CHAIN 17 182 PEPTIDOGLYCAN RECOGNITION PROTEIN.
FT DISULFID 18 140 POTENTIAL.
FT DISULFID 54 60

SO SEQUENCE 182 AA; 20572 MW; 56631E762AE34794 CRC64;

Query Match 37.0%; Score 388.5; DB 1; Length 182;
Best Local Similarity 44.1%; Pred. No. 2.3e-30;
Matches 71; Conservative 35; Mismatches 34; Indels 1; Gaps 1;

OY 28 IVPNEMKALASECAOHLSLPLRYVVVSHTAGSCNTPASCCOQARNOVHYHKTGMCD 87
DB 20 VYTKEMWGLPPIHVEIARPELVIIQHTVSTCNTDAQAQIYVNISSYHNDNINMYD 79

OY 88 VGINFLIGEDGLVYGRGCVTFGANSGLHMPMSIGISWNGNMDRVPTPOAIRAAGLL 147
DB 80 IGSSFIIGNGVYEGAGALHVAHT-YGYNKRKSGITFIYIGNYNDKPKOKSIDALRALI 138

OY 148 ACQVAGALRSNYVLKGRDVOFTLSPGNOLYHLIONMPHYR 188
DB 139 RCGVERGHILTANTHIVGHQOLISTSPGRKLYNEIRMDHF 179

RESULT 4

PGRP_BOMMO STANDARD; PRT; 196 AA.

AC 09XTNO; 09TWD4; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Peptidoglycan recognition protein precursor.

OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Ditrysia; Bombycoidea; Bombycidae; Bombyx.
OX NCBI_Taxid=7091;

RP SEQUENCE FROM N.A., SEQUENCE OF 24-72; 99-118; 147-159 AND 183-196,
TISSUE SPECIFICITY, AND INDUCTION.

RC STRAIN-Kinshu x Showa; TISSUE-Fat body;
RX MEDLINE-99223509; PubMed-10207004;
RA Ochai M., Ashida M.;
RT "A pattern recognition protein for peptidoglycan. Cloning the cDNA and
the gene of the silkworm, Bombyx mori."
RL J. Biol. Chem. 274:11854-11858(1999).

RP SEQUENCE OF 24-43, FUNCTION, AND SUBUNIT.
RC TISSUE-Hemolymph;
RX MEDLINE-96278824; PubMed-8662762;

RA Yoshida H., Kinoshita K., Ashida M.;
RT "Purification of a peptidoglycan recognition protein from hemolymph of
the silkworm, Bombyx mori."
RL J. Biol. Chem. 271:13854-13860(1996).

CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND TRIGGERS THE
PROPODOLIXIDASE CASCADE WHICH IS AN IMPORTANT INSECT DEFENSE
MECHANISM.

CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN FAT BODY,
EPITHELIAL CELLS AND HEMOCYTES. NOT DETECTED IN MALPIGHIAN
TUBULES, SILK GLAND OR MIDGUT.

CC -1- INDUCTION: BY BACTERIAL CHALLENGE.
CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AB016605; BAA77210.1; -
DR EMBL; AB016249; BAA77209.1; -
DR HSP; P00806; ILBA.

KW Immune response; Signal.

FT SIGNAL 1 23
FT CHAIN 24 196 PEPTIDOGLYCAN RECOGNITION PROTEIN.
FT DISULFID 25 147 POTENTIAL.
FT DISULFID 61 67

SO SEQUENCE 196 AA; 21626 MW; 225AD76EC24DA900 CRC64;

Query Match 32.9%; Score 345.5; DB 1; Length 196;
Best Local Similarity 38.3%; Pred. No. 3.4e-26;
Matches 72; Conservative 36; Mismatches 71; Indels 9; Gaps 3;

OY 1 MLIALMLPSILRLGAAOETDPACCSPIVRNEMKALASECAOHLSLPLRYVVVSHTAGS 60
DB 8 VYALALSLI-----TEIADCD-VVSKKQMDGLIPHVSYLARPVSLIVQHTVTP 59

OY 61 SCNTPASCCOQARNOVHYHKTGMCDVGNFLIGEDGLVYBERGNVTFGASHLNNPM 120
DB 60 FCRFDAGCELVNRIGQNHQALQYWDIGSPVVGKGVYEGSGWLHVAHT-YGYNSR 118

OY 121 SIGISFMGNMDRVPTPOAIRAAGLLACGAVAGALRSNYVLKGRDVOFTLSPGNOLYH 180
DB 121 SIGISFMGNMDRVPTPOAIRAAGLLACGAVAGALRSNYVLKGRDVOFTLSPGNOLYH 180


```

05 Pseudomonas sp. (strain ADP).
06 plasmid pADP-1.
0C Bacteria; Proteobacteria.
0X NCBI_TaxID=47660;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97208208; PubMed=9055410;
RA Boundy-Mills K.L., de Souza M.L., Mandelbaum R.T., Wackett L.P.,
RA Sadowsky M.J.;
RT "The atzB gene of Pseudomonas sp. strain ADP encodes the second
RT enzyme of a novel atrazine degradation pathway.";
RL Appl. Environ. Microbiol. 63:916-923(1997).
CC -1- FUNCTION: TRANSFORMS HYDROXYATRAZINE TO N-ISOPROPYLLAMMELIDE
CC (DIHYDROXY-ISOPROPYL-ATRAZINE).
CC -1- CATALYTIC ACTIVITY: 4-(ethylamino)-2-hydroxy-6-(isopropylamino)-
CC 1,3,5-triazine + H2O = N-Isopropylammelide + ethylamine.
CC -1- PATHWAY: Atrazine degradation; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE ATZ/TRZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D66917; AAC45138.1; -.
DR InterPro: IPR002604; ATZ_TRZ.
DR Pfam: PF01685; ATZ_TRZ; 1.
KW Hydrolase; Plasmid.
SQ SEQUENCE 481 AA; 5213 MW; BF2EF1C5891CDA41 CRC64;

Query Match 8.28; Score 86.5; DB 1; Length 481;
Best Local Similarity 23.68; Pred. No. 0.72;
Matches 33; Conservative 20; Mismatches 62; Indels 25; Gaps 5;

QY 26 SPIVPRNEMKALASECAQHLSPRLRYVVSHTTASSCNTPTASCOQARNOYHNKTLGW 85
Db 219 SPIVLAYPEPTEFESARLARLHGLVSL-----HTHNGEGETPRM-----VARFERSLDW 265
QY 86 CDVGTNPLIGSDGLYBERGKNTFGASHGLHNPMSIGISMGVYMDRVPPQAIRAAGC 145
Db 266 CE--NRGFGVGD--VWLAHGEFTPAADIALRL-----AATGTGVAHCPAPVPLVGAEV 313
QY 146 LLACGVAAGALRSNYVLKGH 165
Db 314 TDIPMAAAGVRGFGVGDGH 333

RESULT 8
SURE_HELPJ STANDARD; PRT; 267 AA.
ID SURE_HELPJ
CC 092KSD;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acid phosphatase sure (EC 3.1.3.2).
DE SURE GN JHP0865.
OS Helicobacter pylori j99 (Campylobacter pylori j99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=65963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,
RA Tummala P.U., Caruso A., Girtle-Nickelsen M., Mills D.W., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human

```

```

RT gastric pathogen Helicobacter pylori.
RL Nature 397:176-180(1999).
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -1- COFACTOR: Magnesium (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE SURE ACID PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001516; AAD06448.1; -.
CC HSSP: P96112; 1J9L.
CC InterPro: IPR002828; SURE.
CC Pfam: PF01975; Sure; 1.
CC ProDom: PD005378; Sure; 1.
CC TIGRFAMs: TIGR00087; sure; 1.
CC KM Hydroxylase; Magnesium; Complete proteome.
CC FT METAL 9 MAGNESIUM (BY SIMILARITY).
CC FT METAL 10 10 MAGNESIUM (BY SIMILARITY).
CC FT METAL 40 40 MAGNESIUM (BY SIMILARITY).
CC FT METAL 97 97 MAGNESIUM (BY SIMILARITY).
CC FT ACT SITE 129 129 POTENTIAL.
CC SO SEQUENCE 267 AA; 28994 MW; 1214D2367FB5C8B8 CRC64;

Query Match 7.6%; Score 79.5; DB 1; Length 267;
Best Local Similarity 26.1%; Pred. No. 1.7;
Matches 41; Conservative 26; Mismatches 73; Indels 17; Gaps 9;

QY 17 QETEDPACSPPIVRNEMKALASCAQHLSPRLRYVVSHTAGS-----SCNTPASCQQA 72
DB 22 QALEMAETIYYVARKHE-KSACSCIT-ITAPLRKKEKGRHRYRIDGTPSCVYLA 79
QY 73 RNVOHYHM---KTLGMCDCVGNFLIGEDGLYEGRGWTFGASHGHLMPNSIGISFMGN 129
DB 80 INELFKHYVCFDVLVSGINLGSN--MGED-TYSG---TVAGAIESTIGVPSIAISQILS 133
QY 130 YMDRPTPTQAIRAAGLACGVAAGALSRNRYLKGR 166
DB 134 NKNK-NTPLSFDLAKITQ-DLVONVFTKGYPLKGRK 168

RESULT 9
SURE_HELPY STANDARD; PRT; 267 AA.
AC 025584;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acid phosphatase sure (EC 3.1.3.2).
GN SURE OR HP0930.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_Taxid=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Loftus B., Richardson D., Dodson R., Kirkness E.F., Peterson S.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;

```

```

RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -1- COFACTOR: Magnesium (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE SURE ACID PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000602; AAD07975.1; -.
CC HSSP: P96112; 1J9L.
CC InterPro: IPR002828; SURE.
CC Pfam: PF01975; Sure; 1.
CC ProDom: PD005378; Sure; 1.
CC TIGRFAMs: TIGR00087; sure; 1.
CC KM Hydroxylase; Magnesium; Complete proteome.
CC FT METAL 9 MAGNESIUM (BY SIMILARITY).
CC FT METAL 10 10 MAGNESIUM (BY SIMILARITY).
CC FT METAL 40 40 MAGNESIUM (BY SIMILARITY).
CC FT METAL 97 97 MAGNESIUM (BY SIMILARITY).
CC FT ACT SITE 129 129 POTENTIAL.
CC SO SEQUENCE 267 AA; 30024 MW; 2477244D1F780D62 CRC64;

Query Match 7.6%; Score 79.5; DB 1; Length 267;
Best Local Similarity 26.1%; Pred. No. 1.7;
Matches 41; Conservative 26; Mismatches 73; Indels 17; Gaps 9;

QY 17 QETEDPACSPPIVRNEMKALASCAQHLSPRLRYVVSHTAGS-----SCNTPASCQQA 72
DB 22 QALEMAETIYYVARKHE-KSACSCIT-ITAPLRKKEKGRHRYRIDGTPSCVYLA 79
QY 73 RNVOHYHM---KTLGMCDCVGNFLIGEDGLYEGRGWTFGASHGHLMPNSIGISFMGN 129
DB 80 INELFKHYVCFDVLVSGINLGSN--MGED-TYSG---TVAGAIESTIGVPSIAISQILS 133
QY 130 YMDRPTPTQAIRAAGLACGVAAGALSRNRYLKGR 166
DB 134 NKNK-NTPLSFDLAKITQ-DLVONVFTKGYPLKGRK 168

RESULT 10
YGIU_ECOLI STANDARD; PRT; 356 AA.
AC P42591;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ygiu precursor.
GN YGIU OR B3079.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U18997; AAA57880.1; -
DR EMBL: AE000390; AAC7611.1; -
DR Ecogene: EG12721; y9jJ.

DR Hypothetical protein; signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 356

FT POTENTIAL.
FT HYPOTHELICAL PROTEIN YGJJ.
SQ SEQUENCE 356 AA: 40099 MW: 51D6089E9284385 CRC64;

Query Match
Best Local Similarity 7.4%; Score 77.5; DB 1; Length 356;
Matches 46; Conservative 17; Mismatches 76; Indels 77; Gaps 11;

OY 35 KALASECAQHLSPRYVYVSHTAGSS-----NTPASCOQQAQRYVHYMKTLGMC 86

DB 2 KLITPCCRLALPPCYAF--SAGEEARPAHDDTKTPTATSTSPSPFRFGELGVG-- 57

OY 87 DVGYNFLIGED-----GLYEGRGWNTGAHSG-----HLWN 118

DB 58 --GYMDLEGENHKKYSDGYIEGLEMKYGSWGLY--GEGWTYQVDHGNMVPVPSHG 114

OY 119 PMSIIS-FMGYMDRVPVPPQAIRAAGILACGVAAGAIRS-----NYLAKGR 166

DB 115 GFEIGINREYGGYRNDGTETIMLSLQ-----DSSIDDLQWMDFTPDGLYVLPNTR 166

OY 167 DVQRTLS-----PGNOLYHLI-----QNMPPH 188

DB 167 DMTALKYONLSGNERVSYATPAGHDESKAMLP 202

RESULT 11

FBL4_CRIGR STANDARD; PRT; 443 AA.

AC 055058; 2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE EGF-containing fibulin-like extracellular matrix protein 2 precursor

GN (Fibulin-4) (FBL-4) (H411 protein).

OS Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Cricetus.

OX NCBI_TaxID=10029;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Ovary;

RA Heine H., Delude R.L., Monks B., Golenbock D.F.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR pfam: PF00008; EGF_4.
DR SMART: SM00179; EGF_CA; 4.
DR SMART: SM00001; EGF_Like; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 4.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 4.
DR PROSITE: PS01187; EGF_CA; 6.

KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.

FT SIGNAL 1 25
FT CHAIN 26 443

FT POTENTIAL.
FT EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR

FT MATRIX PROTEIN 2.

FT EGF-LIKE 1, DIVERGENT.
FT EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

FT EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).

FT EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).

FT EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).

FT EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

Query Match
Best Local Similarity 7.2%; Score 75.5; DB 1; Length 443;
Matches 24; Conservative 11; Mismatches 26; Indels 19; Gaps 4;

OY 2 LLMALPSSLRLGAA--QETEDPACCSPIVPRNEMKALASECAQHLSPRYVYVSHTAG 59

DB 12 LLMAL-LLLLLGASRPDSPEPDSYTCEDYEMDASQHCND-----V 55

OY 60 SSCNT-PASCOQQAQRYVHY 78

DB 56 NECLTPEACKGEMKCINHY 75

RESULT 12

VS41_GIALA STANDARD; PRT; 687 AA.

AC P92127;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Variant-specific surface protein VSP41 precursor (CRISP-90).

OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.

OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.

OX NCBI_TaxID=5741;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=02-4A1.

RA Papantastasiou P., Bruderer T., Li Y., Bonnell C., Koehler P.;

RL Mol. Biochem. Parasitol. 86:13-27(1997).

CC CHARACTERIZATION: PubMed-9078242;
CC MEDLINE-97333006; McConville M.J., Ralton J., Koehler P.;

CC The variant-specific surface protein of Giardia, VSP41, is a

RT glycosylated and palmitoylated protein.";
 RL Biochem. J. 322:49-56(1997).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, ANCHORED TO THE
 CC PLASMA MEMBRANE.
 CC -1- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
 CC AT THE REDUCING TERMINUS.
 CC -1- PTM: PALMITOYLATED.
 CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z83743; CAB06038.1; -.
 DR HSSP: O14763; ID06.
 DR GlycoSuiteDB: P92127; -.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR005127; Giardia_VSP.
 DR Pfam: PF03302; VSP; 2.
 DR SMART: SM00181; EGF; 2.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00261; Fu; 3.
 DR Antigen: Repeat, Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
 KW Signal.
 FT SIGNAL 1 14 POTENTIAL.
 FT CHAIN 15 687 VARIANT-SPECIFIC SURFACE PROTEIN VSP4A1.
 FT DOMAIN 15 660 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 661 681 POTENTIAL.
 FT DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;
 Query Match 7.1%; Score 75; DB 1; Length 687;
 Best Local Similarity 22.4%; Pred. No. 14;
 Matches 46; Conservative 19; Mismatches 72; Indels 68; Gaps 12;
 QY 14 GAQEDPDA-----C-CPPIPVNEMKALASBQAHLSPLRVYVSHTGSSCNTPAS 67
 DB 403 GCABCTREPGLKCTCKP---NRKPAGTSD-----NTTCREKT---CEDTV 445
 QY 68 CQO-----QARNVQHYHMKTLGMDVGYNFLIGEDGLVEGRGWNTGAHSHLM 117
 DB 446 CGGTSGACDAIVIDANGKEHY---CSYCGETNKPFI--DGLCTDNKGTN--ACCTDHTC 498
 QY 118 NPASIG-ISTMGN--YKDRP-----TPQAIRAAQGLIAC 149
 DB 499 SYCAAGFFLWGGCYKIDTVPGSYMCKSTYAGVCDTPNANRRFVVPKAIASQSYLAC 558
 QY 150 GVAQALRSNYVLKGRHDVORTLSP 174
 DB 559 GNPGLTIGAGNAIVGVGCSQCTAP 583
 RESULT 13
 GLGB STRAU STANDARD; PRT; 764 AA.
 AC P52980;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (Glycogen branching
 DE enzyme).
 GN GLGB.
 OS Streptomyces aureofaciens.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1894;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 10762 / CCM 3239;
 RX MEDLINE=94347823; PubMed=8068720;
 RA Honerova D., Kormanec J.;
 RT "Cloning of the putative glycogen branching enzyme gene, glgB, from
 RT Streptomyces aureofaciens.";
 RL Biochim. Biophys. Acta 1200:334-336(1994).
 CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
 CC glycogen.
 CC -1- PATHWAY: Glycogen biosynthesis; third step.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L11647; AAA67437.1; -.
 DR InterPro: IPR000461; Alpha_amylase.
 DR InterPro: IPR004193; Isoamylase_N.
 DR Pfam: PF00128; alpha-amylase; 1.
 DR Pfam: PF02922; Isoamylase_N; 1.
 KW Glycogen biosynthesis; Transferase; Glycosyltransferase.
 FT ACT_SITE 440 440 BY SIMILARITY.
 FT ACT_SITE 493 493 BY SIMILARITY.
 FT ACT_SITE 561 561 BY SIMILARITY.
 SQ SEQUENCE 764 AA; 85325 MW; 6B45482E4A268ACF CRC64;
 Query Match 7.1%; Score 75; DB 1; Length 764;
 Best Local Similarity 32.6%; Pred. No. 15;
 Matches 29; Conservative 8; Mismatches 30; Indels 22; Gaps 6;
 QY 90 YNFL-IGEDLVEYGRS-----NMFPGA---HSG-----HLNPNISGISFMG--NY 130
 DB 137 YRFLPALGELDLHLIGEBRHEMLTALSGPMEHGCVAGTRFTVAPNALGVRVYGDPSY 196
 QY 131 MDRVPTPQAIRAAGC---LLACGVAQAL 156
 DB 197 WDVAAYPMRSLGASGWELFLPGVAEGAL 225
 RESULT 14
 MUDD_CHLTR STANDARD; PRT; 803 AA.
 ID MUDD_CHLTR
 AC O84767;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE MURC/del bifunctional enzyme [includes: UDP-N-acetylmuramate--alanine
 DE ligase (EC 6.3.2.8) (UDP-N-acetylmuramoyl-L-alanine synthetase); D-
 DE alanine--D-alanine ligase (EC 6.3.2.4) (D-alanylalanine synthetase)]
 DE (D-Ala-D-Ala ligase)].
 GN MURC/DDL OR CT762.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/DW-3/Cx;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- FUNCTION: CELL WALL FORMATION.
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl + L-alanine - ADP +
 CC phosphate + UDP-N-acetylmuramoyl-L-alanine.

```

CC -1- CATALYTIC ACTIVITY: ATP + D-alanine + D-alanine -> ADP + phosphate
CC + D-alanyl-D-alanine.
CC -1- PATHWAY: Peptidoglycan biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE MURCEDE
CC FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE D-ALANINE-
CC D-ALANINE LIGASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001348; AAC68357.1; -.
CC HSSP: P07862; 110M.
CC InterPro: IPR000291; Data_Data_Ligase.
CC InterPro: IPR000713; Mur_Ligase.
CC InterPro: IPR004101; Mur_Ligase_C.
CC Pfam: PF01225; Mur_Ligase; 1.
CC Pfam: PF01820; Data_Data_Ligase; 1.
CC Pfam: PF02875; Mur_Ligase_C; 1.
CC TIGRPFAMs: TIGR01082; murC; 1.
CC TIGRPFAMs: TIGR01205; Data_Data_Ligase; 1.
CC PROSITE: PS00843; DATA_DATA_LIGASE_1; 1.
CC PROSITE: PS00844; DATA_DATA_LIGASE_2; FALSE_NBS.
CC Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
CC ATP-binding; Multifunctional enzyme; Complete proteome.
CC FT DOMAIN 1 446 UDP-N-ACETYLTRANSFERASE--ALANINE LIGASE.
CC FT NP_BIND 447 803 ATP (POTENTIAL).
CC FT NP_BIND 111 117 ATP (POTENTIAL).
CC SQ SEQUENCE 803 AA; 89230 MW; 82B8523BF1F4C1C8 CRC64;

Query Match 7.1%; Score 75; DB 1; Length 803;
Best Local Similarity 25.0%; Pred. No. 16;
Matches 29; Conservative 24; Mismatches 41; Indels 22; Gaps 7;

QY 86 CDVGVNFLIGEDG-----LVYEGRWNV-FTGAHSHLMNPMISGIFMGNYMDRVPPTP 137
DB 220 CLOGHTF--GIDSSCDLHILSYQEGWRLYFTAKYQDVYADIEQVLGMHVNLA--- 274

QY 138 QAIRAAGL-LACGVAGALNSVYKGRDYOQLTSPGN--QLXHLIOMPHYKYS 190
DB 275 ----AAGIALSLGIDEGAIRNAF--RGFSGVORLORKNSSEPLFLEDYAHPS 324

RESULT 15
TS4_GIALA STANDARD; PRT; 713 AA.
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major surface-labeled trophozoite antigen 417 precursor.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 30957 / WB;
RX MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Alez S.B., Reiner D.S.,
RA Mc Caffery M., So M., Guiney D.G.;
RA "Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia.",
RL Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RC STRAIN=AD-1;

```

```

RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis."
RL Gene 129:257-262(1993).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -1- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M33641; AAA02688.1; -.
CC EMBL: M97488; AAA02581.1; -.
CC PIR: A35502; A35502.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR005127; Giardia_VSP.
CC Pfam: PF03302; VSP; 2.
CC SMART: SM00181; EGF; 1.
CC SMART: SM00001; EGF-like; 1.
CC SMART: SM00261; FU; 3.
CC Signal: Antigen; Glycoprotein; Transmembrane; Repeat.
CC FT CHAIN 1 17
CC FT 18 713 MAJOR SURFACE-LABELLED TROPHOZOITE
CC FT 18 713 ANTIGEN 417.
CC FT TRANSMEM 680 708 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 709 713 POTENTIAL.
CC FT CARBOHYD 289 289 CYTOPLASMIC (POTENTIAL).
CC FT CARBOHYD 676 676 N-LINKED (GLCNC. . .) (POTENTIAL).
CC FT VARIANT 582 582 A -> T (IN STRAIN ADELAIDE-1).
CC FT VARIANT 606 606 A -> S (IN STRAIN ADELAIDE-1).
CC SQ SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;

Query Match 7.0%; Score 74; DB 1; Length 713;
Best Local Similarity 22.6%; Pred. No. 18;
Matches 45; Conservative 19; Mismatches 85; Indels 50; Gaps 9;

QY 2 LIALALPSLLRLGAQETEDPAC--CSPYPRNEMKALASECAQHLSPLRYVVSHTAG 59
DB 6 LLAIVYLQIARFACIQEADDGKCKTGCTIGDTW---CSECGANAYAPVNGOCVDVNAE 62

QY 60 SSCNTPASCCQARNVQHYHMKTLGWC-DVGYNFLIGEDGLVYEGRWNTGAHSHLMN 118
DB 63 GPSKT--LCPOH-----SAGKCTCGGNSFWKDCYSSGREL-----PGH--- 101

QY 119 PMSIGSFNGNWDVRYPTP-----QAIRAGCLLACGVAGALNSVYKGRDYOQLT 172
DB 102 --SLCISSDGDGYCTAAGGYFAPVPGANTQESVYIACGDTGV-----TI 144

QY 173 SPGNOLYHLIOMPHYRSP 191
DB 145 AAGGNTRYKGIADCAECSSAP 163

```

Search completed: November 12, 2002, 02:23:54
 Job time : 87 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 01:41:04 ; Search time 144 Seconds

(without alignments)
273.299 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAMALPSILRLGAAQETE.....LSPGNQLYHLIQQNPHYRSP 191

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	772	73.5	193	6 Q9GK12	Q9GK12 camelus dro
2	717	68.2	190	6 Q8SP7	Q8SP7 bos taurus
3	667.5	63.5	183	11 Q9JLN4	Q9JLN4 rattus norv
4	441	42.0	341	4 Q96L89	Q96L89 homo sapien
5	414.5	39.4	184	5 Q9V4X2	Q9V4X2 drosophila
6	413	39.3	373	4 Q96L88	Q96L88 homo sapien
7	405.3	38.6	185	5 Q9V3B7	Q9V3B7 drosophila
8	403.5	38.4	185	5 Q95S09	Q95S09 drosophila
9	386	36.7	86	6 Q8WME9	Q8WME9 sus scrofa
10	379	36.1	203	5 Q9VYX7	Q9VYX7 drosophila
11	371.5	35.3	368	4 Q9HD75	Q9HD75 homo sapien
12	369	35.1	195	5 Q8HSZ1	Q8HSZ1 bombyx mori
13	353.5	33.6	500	5 Q8T5Q2	Q8T5Q2 drosophila
14	348.5	33.2	576	4 Q96PD5	Q96PD5 homo sapien
15	347.5	33.1	634	4 Q96N74	Q96N74 homo sapien
16	347	33.0	345	5 Q9VXN9	Q9VXN9 drosophila

17	339	32.3	194	5 Q9BL11	Q9BL11 bombyx mori
18	336.5	32.0	208	5 Q9BL12	Q9BL12 bombyx mori
19	334	31.8	195	5 Q9V369	Q9V369 calpodes et
20	330	31.4	186	5 Q9VS97	Q9VS97 drosophila
21	329.5	31.4	190	5 Q9VY97	Q9VY97 drosophila
22	329.5	31.4	530	11 Q8VCS0	Q8VCS0 mus musculu
23	328	31.2	529	11 Q9QX22	Q9QX22 mus musculu
24	327.5	31.2	369	5 Q8SXQ7	Q8SXQ7 drosophila
25	327	31.1	337	5 Q8VSW0	Q8VSW0 drosophila
26	317	30.2	215	5 Q9VGN3	Q9VGN3 drosophila
27	315	30.0	500	11 Q9QX21	Q9QX21 mus musculu
28	289.5	27.5	182	5 Q9VY96	Q9VY96 drosophila
29	248.5	23.6	611	5 Q9VSV9	Q9VSV9 drosophila
30	229.5	21.8	280	5 Q9VSV8	Q9VSV8 drosophila
31	229.5	21.8	299	5 Q9GNK7	Q9GNK7 drosophila
32	224	21.3	539	16 Q86334	Q86334 mycobacteri
33	221	21.0	520	5 Q9GNK5	Q9GNK5 drosophila
34	181.5	17.3	138	5 Q9GNK6	Q9GNK6 drosophila
35	177.5	16.9	416	16 Q9KZK5	Q9KZK5 streptomyce
36	151	14.4	222	16 Q8XMX3	Q8XMX3 clostridium
37	148.5	14.1	308	16 Q9S2P9	Q9S2P9 streptomyce
38	147.5	14.0	282	5 Q9GNP7	Q9GNP7 drosophila
39	147.5	14.0	505	5 Q9VRP7	Q9VRP7 drosophila
40	133.5	12.7	304	16 Q8XLA4	Q8XLA4 clostridium
41	126	12.0	151	9 Q9T132	Q9T132 bacteriophag
42	91	8.7	879	10 Q9FZAB	Q9FZAB arabidopsis
43	85.5	8.1	642	13 P79941	P79941 xenopus lae
44	82	7.8	268	5 Q62353	Q62353 caenorhabdi
45	80.5	7.7	281	11 Q88812	Q88812 mus musculu

ALIGNMENTS

RESULT 1

Q9GK12

AC Q9GK12: PRELIMINARY: PRT: 193 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

GN Peptidoglycan recognition protein precursor.

OS Camelus dromedarius (Dromedary) (Arabian camel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.

OX NCBI_TaxID=9838;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LACTATING MAMMARY GLAND;

RA Kappeler S.R., Farah Z., Puhon Z.;

RT "Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan

RL Recognition Protein.";

RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Kappeler S.R., Farah Z., Puhon Z.;

RT "The peptidoglycan recognition protein is expressed in the lactating

RL mammary gland of camels and binds to lactic acid bacteria.";

RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AJ131676; CAC10553.1; -

DR EMBL: AJ409286; CAC84130.1; -

KW Signal.

FT SIGNAL. 1 21

FT CHAIN. 22 193

SO SEQUENCE 193 AA; 21377 MW; B6A1BD818030A7CB CRC64;

Query Match

Best Local Similarity 73.5%; Score 772; DB 6; Length 193;

Matches 140; Conservative 21; Mismatches 27; Indels 2; Gaps 1;

1 MLAMALPSILRLGAAQETEDPACCSPIYPRNWKRLASCAQHLSTPLATVYVSHRAGS 60

[illegible][illegible]

RP SEQUENCE FROM N.A..
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=SPLEEN;
RA Reiman A., Teodoreski E.E., Krueger J.M.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF154114; AAF33252.1; -
SQ SEQUENCE 183 AA; 20590 MW; 5B9CB1B7AA8A2EC21 CRC64;

Query Match 63.5%; Score 667.5; DB 11; Length 183;
Best Local Similarity 65.3%; Pred. NO. 4.4e-60;
Matches 124; Conservative 23; Mismatches 34; Indels 9; Gaps 3

QY	1	MLLAMA-LPSTLRLRGAOETEDPACSCPIVYPRNEMKLAEECAOHSLPRLRYVVSHTAG	59
Db	1	MLFMMAFPFALLGLA-----DBCCF-VYPRSEMKRLBPESCKGLKKPRYYVISHTAG	52
QY	60	SSCNTPASCOOQAAANVOHYHMKTLGMCVDVGYNFLIGEDGLVYEGRGWNTFGASHGLMNP	119
Db	53	SFGSSPDSCEQOAAANVQYLQMKQGLGMCDDVAYNFLIGEDGHVYEGRGWTKGHDHTGPIWNP	112
QY	120	MSIISTEMGNTMDRVPFPQAIRAAGLGLACVAGALRSNTYVLKGHRDVRITLSPGNQLY	179
Db	113	MSIISTEMGDSHYRVPKRALRALNLMLKCGVSEGEFLRSNYEYKGRHDVOSTLSPGQPLY	172
QY	180	HLIIONPBYR	189
Db	173	ELIOSMDHYR	182

RESULT 4			
096Lb9			
ID	096Lb9	PRELIMINARY:	PRT: 341 AA.
AC	096Lb9		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Peptidoglycan recognition protein-I-alpha precursor.		
GN	PGLYRPA1.PHA		
OS	Homo sapiens (Human).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21433985; PubMed=11461926;		
RA	Liu C., Xu Z., Gupta D., Dzialski R.;		
RT	"Peptidoglycan recognition proteins, a novel family of four human		
RT	innate immunity pattern recognition molecules.";		
RL	J. Biol. Chem. 276:34686-34694(2001).		
DR	EMBL; AY035376; AAK72484.1; -.		
FT	Signal.		
KW	Signal.		
FT	1	17	POTENTIAL.
QO	SEQUENCE	341 AA;	37611 MW; 84DD5AA97B632076 CRC64;

[illegible]

```

RESULT 5
Q9VAX2 PRELIMINARY; PRT; 184 AA.
AC Q9VAX2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE CG14745 protein.
GN GPR-SC2 OR CG14745.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Berens P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borokov D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu K., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclio J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003836; AAC59051.1; -.
DR HSSP; P00806; L1BA.
DR FlyBase; FBgn0043575; GPR-SC2.
SQ SEQUENCE 184 AA; 19829 MW; 0F99D04914B07238 CRC64;

Query Match 39.4%; Score 414.5; DB 5; Length 184;
Best Local Similarity 47.2%; Pred. No. 2.5e-34;
Matches 77; Conservative 31; Mismatches 54; Indels 1; Gaps 1;

```

```

QY 148 ACYAGALRSNVYLGKGRDVOFTLSPGNQYLHIONMPHYRS 190
DB 142 SDVSRGQIVSGYLLGHROVGSTECPGTINMEIPTWSMKA 184

RESULT 6
Q96LB8 PRELIMINARY; PRT; 373 AA.
AC Q96LB8;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Peptidoglycan recognition protein-I-beta precursor.
GN PGLYPLBETA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21433985; PubMed=11461926;
RA Liu C., Xu Z., Gupta D., Dzidarski R.;
RT "Peptidoglycan recognition proteins, a novel family of four human
RT innate immunity pattern recognition molecules."
RL J. Biol. Chem. 276:34686-34694(2001).
DR EMBL; AY035377; AA072485.1; -.
DR InterPro; IPR020086; Aldehyde_dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN.1.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 373 AA; 40663 MW; 148BA166018A66AA CRC64;

Query Match 39.3%; Score 413; DB 4; Length 373;
Best Local Similarity 40.0%; Pred. No. 8.8e-34;
Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

```

```

QY 4 ANAPSLRL---LGAQETEDPACCSPIYPRNEMKALASECAHSLPRLYVVSITAG 59
DB 185 SYVQPLVKGENCELAPROKTSLKACPGVPRSRVWCAREFHCBP-MTPAKYGIITTAG 243
QY 60 SSCNTPASCOQOARNOYHMKTLGMCDDVGVNVLIGDGLVYGRGNFGAHSGLMNP 119
DB 244 RITCNIDECRLVDRDQSFIDRLKSCDIGNLVGDGAYIEGVGNVGS-STPRTD 302
QY 120 MSIGISFMGNMVRVTPQAIRAAGLLACGVAAGALRSNVYLGKGRDVOFTLSPGNQYL 179
DB 303 IALGIFFMGFTGCIIPNAALAEAAODLICAMVKGITLPYLLVGHSDVARTLSPGQALY 362
QY 180 HLIQNPYHR 189
DB 363 NIISTWPHRK 372

RESULT 7
Q9V3B7 PRELIMINARY; PRT; 185 AA.
AC Q9V3B7;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE CG8577 protein and CG14746 protein (Peptidoglycan-recognition protein-
DE SC1B).
GN GPR-SC1A OR CG8577.3 OR CG14746.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;

```


OY 118 NPMISIGISFMGNMDRVPPOAIRAA 143
 DB 61 NPLSLGISFMONKMRVPPAIRRAA 86

RESULT 10

O9VXX7 PRELIMINARY; PRT: 203 AA.
 AC O9VXX7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CG11709 protein (Peptidoglycan-recognition protein-Sa)
 DE (A130827b).
 GN pGRP-SA OR CG11709.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RX MEDLINE=20196006; Pubmed=10731132;
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gockayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DP CL CN BW;
 RX MEDLINE=20558582; Pubmed=11106397;
 RT Werner T., Liu G., Kung D., Ekengren S., Steiner H., Hultmark D.,
 "A family of peptidoglycan recognition proteins in the fruit fly
 Drosophila melanogaster."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).
 RN [3]
 RP SEQUENCE FROM N.A.

RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Drensek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nuno J., Paclet J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE003486; AAF48056.1;
 DR EMBL: AF207541; AAG23735.1;
 DR EMBL: AF207540; AAG23734.1;
 DR EMBL: AY075293; AAL68160.1;
 DR FLYBASE: FBgn030310; pGRP-SA.
 SQ SEQUENCE 203 AA; 22260 MW; D200A6EA79C66731 CRC64;

Query Match 36.1%; Score 379; DB 5; Length 203;
 Best Local Similarity 40.1%; Pred. No. 1.2e-30;
 Matches 76; Conservative 34; Mismatches 76; Indels 2; Gaps 2;

OY 3 LAMALPSILRLGAQETEDPACCSPVPRNEMKALASECAQHSLPLRYVVSHTAGSSC 62
 DB 16 LVLLALFVSACKSRQ-RSPANCPTIKLRQWGGKPSLGLHVOVRPRVYVHHVTVGEC 74
 OY 63 NTPASCOQCARVVOHYHMKTLGCDVGNFLIGEDLVYEGRGNFTGASHGLAMPMSI 122
 DB 75 SLLKCAELIOMQAAVHOMELDFNDISYFLIGNDSIVYEGTGWGLRGANT-YGVNAGT 133
 OY 123 GISFMGNMDRVPPOAIRAAQGLACGYAAGALNSNYVLRGRDVOFTLSPGNQYHLI 182
 DB 134 GIAFGNFEVDKLPSPALQAARDLACGVQGESLSEVDYALIGSQYISQSPCLTYLNET 193
 OY 183 QNMPHYRS 190
 DB 194 QEMPHRLS 201

RESULT 11

O9HD75 PRELIMINARY; PRT: 368 AA.
 AC O9HD75;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Hypothetical 40.0 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang W., Wan T., Cao X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF242518; AAF95959.1;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 368 AA; 40020 MW; 1E74970732A5DAFD CRC64;

Query Match 35.3%; Score 371.5; DB 4; Length 368;
 Best Local Similarity 44.1%; Pred. No. 1.5e-29;
 Matches 64; Conservative 33; Mismatches 47; Indels 1; Gaps 1;

OY 45 LSLPLRYVVSHTAGSSCTPASCQQAQANVOHYHMKTLGCDVGNFLIGEDLVYEGR 104
 DB 224 MTLPAKYGIIHTAGRTCTISDECRILVDISFYIDRLKSCDIGVFLVGQGAIVEGV 283
 OY 105 GNFTGASHGLAMPMSISIFMGNYMDRVPPOAIRAAQGLACGYAAGALNSNYVLC 164
 DB 284 GNNVGS-STPEYDIALAIIITMGFTTGIPNAAALEAADLIQCAVMAGYLTPLYLVG 342
 OY 165 HRDVOFTLSPGNQYHLIOMPHYR 189
 DB 343 HSDVARTLSPGOALYNIISTWPHK 367

RESULT 12

08MSZ1 PRELIMINARY; PRT; 195 AA.

AC 08MSZ1
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Peptidoglycan recognition protein.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Bombycoidea; Bombycidae; Bombyx.
 NCBI_Taxid=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kim S.H., Lee H.S., Kim J.W., Lee Y.S., Ryu K.S., Jun Y., Minoru Y.,
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF441723; AAL32058.1;
 SQ SEQUENCE 195 AA; 21478 MW; 17F85F01BDA035E CRC64;

Query Match 35.1%; Score 369; DB 5; Length 195;
 Best Local Similarity 38.6%; Pred. No. 1.2e-29;
 Matches 76; Conservative 35; Mismatches 66; Indels 20; Gaps 4;

OY 1 MLTAMALPSL-----RLGAOETEDPACCSPVPRNEMKALASECAQHLSPLRY 51
 DB 1 MLVA---PSLLLVFLVSFETLNAASECGE-----IPTEWSTESRRKQPLKSPIDL 50
 OY 52 VVVSHTAGSSCNTPASCOOQARNOVHYHMKTLGMCVDVGNFLIGEDGLYEGRGMTFGA 111
 DB 51 VVIOHTVSNDCFTDEBCLLSVNSLRQHHMLAGFKDGLGSFVAGNGKITYEGAGMNNIGA 110
 OY 112 HSGHLMNPKSIGISFGMNTYDRVTPQATRAAGLACGAAQGLAKSNVYLKGRDYORT 171
 DB 111 HTLH-VNNISIGIGFGDFEKLPTQALQAVDFLACGVENMLTBDYHYVGHQQLINT 169
 OY 172 LSPGNOLYHLIQNMPHY 188
 DB 170 LSPGATVLSIESMPTH 186

RESULT 13

08T502 PRELIMINARY; PRT; 500 AA.

AC 08T502
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Peptidoglycan recognition protein-LC isoform x.
 GN PGRP-LC
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pubmed-11872802;
 RX Choe K.-M., Werner T., Stoven S., Hultmark D., Anderson K.V.,
 RT Requirement for a Peptidoglycan Recognition Protein (PGRP) in Relish
 RL Activation and Antibacterial Immune Responses in Drosophila.
 DR Science 296:359-362(2002).
 DR EMBL; AF500096; AAM18530.1;
 SQ SEQUENCE 500 AA; 54094 MW; A0459DE723A8C720 CRC64;

Query Match 33.6%; Score 353.5; DB 5; Length 500;
 Best Local Similarity 47.2%; Pred. No. 1.5e-27;
 Matches 67; Conservative 22; Mismatches 52; Indels 1; Gaps 1;

OY 47 LPLRYVVSHTAGSSCNTPASCOOQARNOVHYHMKTLGMCVDVGNFLIGEDGLYEGRGW 106
 DB 356 LPLRNVVISHTAEGCESREVCSARVNVVOSFHMDSGMDHIGYVGLVGDGRVYEGRGW 415

OY 107 NFGANSHLMNPKSIGISEMGNMDRVTPQATRAAGLACGAAQGLAKSNVYLKGRH 166
 DB 416 DYVAHTTKG-VNRSGISGIFGTTRKPRNEROLEACOLLQEGVRLKRLTYRLYGRH 474

OY 167 DVORTLSPGNOLYHLIQNMPHY 188
 DB 475 QLSATESPGBELLYIKKMPHY 496

RESULT 14

096PD5 PRELIMINARY; PRT; 576 AA.

AC 096PD5
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Peptidoglycan recognition protein L precursor.
 GN pLtyrP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-21433985; Pubmed-11461926;
 RA Liu C., Xu Z., Gupta D., Dzierski R.;
 RT Peptidoglycan recognition proteins: a novel family of four human
 RT innate immunity pattern recognition molecules.";
 RL J. Biol. Chem. 276:34686-34694(2001).
 DR EMBL; AF384856; AAL05629.1;
 KM Signal.

FT SIGNAL 1
 SQ SEQUENCE 576 AA; 62216 MW; 73EA8713DC54F85A CRC64;

Query Match 33.2%; Score 348.5; DB 4; Length 576;
 Best Local Similarity 41.4%; Pred. No. 5.6e-27;
 Matches 70; Conservative 26; Mismatches 66; Indels 7; Gaps 5;

OY 25 CSPVPRNEMKALASE-CAQHLSPRLRYVVSHT--AGSSCNTPASCOOQARNOVHYHMK 81
 DB 379 CPAIHPRCRMGADYRGRPKLLQPLGLYVHTYVPAAPCTDFTRCANRRSMGRYHQD 438
 OY 82 TLGWCVDVGNFLIGEDGLYEGRGMTFGAHS-GHLMNPKSIGISFGMNTYDRVTPQAT 140
 DB 439 TQGGDGGISFVVSDDYVEGGRMHVGAHTLGH--NSRGEVAIVGNTAALPTEAL 496
 OY 141 RAAQGL-ACGVAAGALRSNVYLKGRDYORTLSPGNOLYHLIQNMPHY 188
 DB 497 RTVADTLPSCAVRAGLPRDIALGHRQLVNTDPCGDALFDLKRMPHF 545

RESULT 15

096N74 PRELIMINARY; PRT; 634 AA.

AC 096N74
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CNNA FLJ31320 f1s, clone LIVER100542, moderately similar to Mus
 DE musculus TAGL-alpha mRNA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Maeda-tsuma M., Murakami K., Kanehori K., Takahashi-pujii A., Oshima A.
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,
RA Masuno Y., Nagai K., Isonai T.,
RT "NEBD human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL, AK055882; BAB71034.1;
SQ SEQUENCE 634 AA; 67970 MW; 93E2032F3CE3BE70 CRC64;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 00:45:49 ; Search time 143 Seconds

(without alignments)
177.978 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051
Sequence: 1 MLTAWLPRLRLGAQETE.....LSPGNOLYHLIQWPHYRSP 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:.*
1: /SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/genesep-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1051	100.0	191	20	AA00771	Human tag7 clone p
2	1051	100.0	196	21	AA254022	Human PRO1269 prot
3	1051	100.0	196	21	AA25583	Htag7 Protein enco
4	1051	100.0	196	21	AA99400	Chondrosarcoma pep
5	1051	100.0	196	21	AA99400	Human PRO1269 (UNG
6	1051	100.0	196	22	AA66149	Protein of the inv
7	717	68.2	190	18	AAW23722	Bovine granulocyte
8	692	65.8	182	20	AA00770	Mouse tag7 clone p
9	661.5	62.9	181	18	AAW23723	Murine granulocyte
10	494	47.0	116	21	AA64935	Human 5' EST relat

11	414.5	39.4	184	22	ABB69757	Drosophila melanog
12	413	39.3	241	22	ABB53271	Human polypeptide
13	413	39.3	369	22	AAE00693	Human full length
14	413	39.3	369	22	ABB53272	Human full length
15	413	39.3	369	22	AAE00692	Human full length
16	405.5	38.6	185	22	AAE59234	Drosophila melanog
17	405.5	38.6	185	22	ABB69758	Drosophila melanog
18	379	36.1	203	22	ABB70267	Drosophila melanog
19	371.5	35.3	368	21	AA96963	Wound healing tiss
20	348.5	33.2	363	21	AA94863	Human protein clon
21	348.5	33.2	576	22	AA65916	Amino acid sequenc
22	348.5	33.2	634	22	AA65915	Amino acid sequenc
23	347.5	33.1	196	19	AAW37837	Amino acid sequenc
24	347.5	33.1	196	19	AAW37835	Amino acid sequenc
25	347	33.0	345	22	ABB64149	Drosophila melanog
26	332.5	31.6	530	22	AA972664	Murine peptidoglyc
27	330	31.4	186	22	ABB61200	Drosophila melanog
28	329.5	31.4	190	22	ABB64581	Drosophila melanog
29	328.5	31.3	173	19	AAW37834	Recombinant peptid
30	328.5	31.3	173	19	AAW37836	Amino acid sequenc
31	327	31.1	337	22	ABB60644	Drosophila melanog
32	317	30.2	215	22	ABB65711	Drosophila melanog
33	316.5	30.1	243	21	AA96962	C glutamicum prote
34	316.5	30.1	244	21	AA976124	Novel human diago
35	289.5	27.5	182	22	AAE00694	Human peptidoglyc
36	289.5	27.5	182	22	ABB64595	Human zcpai domain
37	228.5	21.8	280	22	ABB60610	Drosophila melanog
38	202	19.2	683	22	ABB60590	Drosophila melanog
39	196.5	18.7	363	22	ABG27581	Drosophila melanog
40	193	18.4	174	22	AA972663	Human peptidoglyc
41	177.5	16.9	132	22	AAE00694	Human zcpai domain
42	147.5	14.0	505	22	ABB61601	Drosophila melanog
43	138	13.1	26	20	AA00774	Human tag7 clone p
44	127.5	12.1	33	22	ABB42485	Peptide #9991 enco
45	127.5	12.1	53	22	ABB25910	Protein #7909 enco

ALIGNMENTS

RESULT 1	AA00771	standard; Protein; 191 AA.
ID	AA00771	
XX	AA00771	
AC	AA00771	
DT	18-MAY-1999	(first entry)
XX	Human tag7 clone protein sequence.	
DE	Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma; melanoma; leukaemia; apoptosis inducer; human.	
KW	melanoma; leukaemia; apoptosis inducer; human.	
KW	melanoma; leukaemia; apoptosis inducer; human.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9902686-A1.	
XX		
PD	21-JAN-1999.	
XX		
PF	10-JUL-1998;	98WO-EP04287.
XX		
PR	11-JUL-1997;	97US-0893764.
XX		
PA	(BOEH) BOEHRINGER INGELHEIM INT. GMBH.	
XX		
PI	Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;	
XX	WPL; 1999-120887/10.	
DR	N-PSDB: AAX21820.	
XX		
XX	New nucleic acid encoding tag7 - used to inhibit tumour growth and	
PT	induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and	
PT	leukaemia	

```

XX Claim 23; Page 126-127; 138pp; English.
PS
XX This sequence is the human tag7 of the invention. Cells containing
CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
CC to produce and purify antibodies; to inhibit growth of mammalian tumours,
CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
CC osteo- or fibro-sarcoma), melanoma or leukaemia; and as a molecular
CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
CC apoptosis. The tag7 coding sequences are also useful as probes for gene
CC mapping and detection of tag7 gene expression, and as primers. Antibodies
CC against tag7 are used as reagents for detecting tag7; as an antagonist of
CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
CC metastasis.
XX
SQ Sequence 191 AA;
Query Match 100.0%; Score 1051; DB 20; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.6e-104;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLAAALPSLLRLGAAGETEDPACCSPIVPRNEMKALASECAOHLSPLRVYVSHTAGS 60
DB 1 MLAAALPSLLRLGAAGETEDPACCSPIVPRNEMKALASECAOHLSPLRVYVSHTAGS 60
QY 61 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLVYEGRGWNTGASHGLMNP 120
DB 61 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLVYEGRGWNTGASHGLMNP 120
QY 121 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNTVYLGKHRDVOFTLSPGNOLYH 180
DB 121 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNTVYLGKHRDVOFTLSPGNOLYH 180
QY 181 LIQNMPHYRSP 191
DB 181 LIQNMPHYRSP 191
RESULT 2
AAB24022
ID AAB24022 standard; Protein; 196 AA.
XX
XX AAB24022;
AC
XX
XX 25-JAN-2001 (first entry)
DT
XX
DE Human PRO1269 protein sequence SEQ ID NO:7.
XX
XX Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW identification; tumorigenesis; anticancer; detection.
XX
XX Homo sapiens.
OS
XX
XX WO200053750-A1.
PN
XX
XX 14-SEP-2000.
PD
XX
XX 02-DEC-1999; 99WO-US28551.
PF
XX
XX 08-MAR-1999; 99WO-US05028.
PR
XX 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
XX (GETH ) GENENTECH INC.
XX
XX Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
XX WPI: 2000-594320/56.
XX
XX N-PSDB; AAC58104.
DR

```

```

XX
XX Antibodies specific for PRO polypeptides; used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression.
XX
XX Claim 61; Fig 4; 226pp; English.
XX
XX The present invention describes an antibody that binds to a human
CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO334; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumors in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumors. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human PRO
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
SQ Sequence 196 AA;
Query Match 100.0%; Score 1051; DB 21; Length 196;
Best Local Similarity 100.0%; Pred. No. 2.7e-104;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLAAALPSLLRLGAAGETEDPACCSPIVPRNEMKALASECAOHLSPLRVYVSHTAGS 60
DB 6 MLAAALPSLLRLGAAGETEDPACCSPIVPRNEMKALASECAOHLSPLRVYVSHTAGS 65
QY 61 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLVYEGRGWNTGASHGLMNP 120
DB 66 SCNTPASCCOQARNVQHYHMKITLGMCDVGYNFLIGEDGLVYEGRGWNTGASHGLMNP 125
QY 121 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNTVYLGKHRDVOFTLSPGNOLYH 180
DB 126 SIGISFMGNMYMDRVPPTPOAIRAAGGLACGVAAGALRSNTVYLGKHRDVOFTLSPGNOLYH 185
QY 181 LIQNMPHYRSP 191
DB 186 LIQNMPHYRSP 196
RESULT 3
AAB25583
ID AAB25583 standard; Protein; 196 AA.
XX
XX AAB25583;
AC
XX
XX 21-NOV-2000 (first entry)
DT
XX
DE Htag7 protein encoded by human secreted protein gene #8.
XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritis;
KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnery; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human.
XX
XX Homo sapiens.
OS
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX
XX

```


OY 1 MLTAMALPSILIRIGAQETEDPACCSPIVPRNEKALASECAOHLSTPIRYVVSHTAGS 60
DB 6 MLTAMALPSILIRIGAQETEDPACCSPIVPRNEKALASECAOHLSTPIRYVVSHTAGS 65
OY 61 SCMTFPAACCOQAAHVQHYHMKTLGWCVDVGYNFLIGEDGLVEGRGMFTGAHSGHLNPM 120
DB 66 SCMTFPAACCOQAAHVQHYHMKTLGWCVDVGYNFLIGEDGLVEGRGMFTGAHSGHLNPM 125
OY 121 SIGISMGNMNDVPPFQALRAAGGLLACGVAAGALNSNYLKGRVQRTLSFGNOLYH 180
DB 126 SIGISMGNMNDVPPFQALRAAGGLLACGVAAGALNSNYLKGRVQRTLSFGNOLYH 185
OY 181 LIONMPTHYRSP 191
DB 186 LIONMPTHYRSP 196
RESULT 5
AAV9400
ID AAV9400 standard; Protein; 196 AA.
XX AAV9400;
AC AAV9400;
DT 08-AUG-2000 (first entry)
XX Human FRO1269 (UNQ639) amino acid sequence SEQ ID NO:216.
DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
XX Transmembrane; secretion; Immunoadhesion; pharmaceutical; screening.
KM Homo sapiens.
XX WO200012708-A2.
PN 09-MAR-2000.
PD 09-MAR-2000.
XX 01-SEP-1999; 99WO-US20111.
PF 01-SEP-1998; 98US-0098716.
XX 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099604.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.

PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.

XX Bovine granulocyte peptide A precursor (antimicrobial BGP-A).
 XX Antimicrobial peptide; antibiotic; antibacterial; antifungal;
 KW fungicide; antiprotoczoa; protozoacide; antiviral; virucide;
 KW bovine granulocyte peptide A; BGP-A; preservative; sepsis;
 KW endotoxaemia; cattle.
 XX Bos taurus.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Sig-peptide
 FT Peptide 22..177
 FT /label= Propeptide
 FT Peptide 178..190
 FT /label= Mat-peptide
 FT /note= "BGP-A antimicrobial peptide (Claim 2)"
 XX
 XX WO9729765-A1.
 XX
 XX 21-AUG-1997.
 XX
 XX 13-FEB-1997; 97WO-US02218.
 XX
 XX 16-FEB-1996; 96US-0011834.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Selsled ME;
 XX
 XX WPI; 1997-424753/39.
 XX N-PSDB; AAT78509.
 XX
 XX Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
 PT useful therapeutically, as preservatives for food, in water
 PT treatment and in agriculture
 XX
 XX Claim 8; Fig 4; 56pp; English.
 XX
 XX This protein comprises the precursor (see AAM23722) of a novel,
 CC claimed antimicrobial peptide from bovine neutrophils, designated
 CC bovine granulocyte peptide A or BGP-A (see AAM23724). Its amino acid
 CC sequence was deduced from a cDNA clone (see AAT78509) obtained from
 CC bovine bone marrow. BGP-A and the murine homologue, MGP-A (see
 CC AAM23725), exhibit activity against Gram-positive and Gram-negative
 CC bacteria, fungi and viruses, specifically *Staphylococcus aureus*,
 CC *Escherichia coli*, *Candida albicans*, *Salmonella typhimurium* and *C.*
 CC *neoformans* (claimed). They can be used in human or veterinary
 CC medicine (particularly to treat disorders associated with
 CC *lipopolysaccharides*, e.g. sepsis and endotoxaemia) or as
 CC preservatives in food products or in water supplies (claimed).
 CC They can also be applied to crops to reduce post-harvest spoilage
 CC or expressed in transgenic plants to increase their disease
 CC resistance. They have low immunogenicity. Carboxamidated
 CC analogues of BGP-A and MGP-A may also be used.
 CC
 XX Sequence 190 AA:
 SQ

Query Match 68.2%; Score 717; DB 18; Length 190;
 Best Local Similarity 70.1%; Pred. No. 1,5e-68;
 Matches 131; Conservative 20; Mismatches 30; Indels 6; Gaps 1;

QY 3 LAMALPSLRIGAADETEPACCSPIVPRNEKALASECAQHLSLPRVYVVSHTAGSSC 62
 DB 8 LAMVLLALGIGLGAAD-----CGSIVSRKWKALSKCSQRLKOPRYIVVSHHTASVC 61
 QY 63 NPASCOQOQARNVQYHMKTLGMCVGYNFLIGEDGLVYEGRGWNETGAHSGHLMNPM 122
 DB 62 NPASCOQOQARNVQYHMKTLGMCVGYNFLIGEDGLVYEGRGWNETGAHSGHLMNPM 121
 QY 123 GISFNGNMDRVPPTQAIRAAOGLLACGVAOGLRSNYVLKGRDVRITSPGNOLYHL 182

DB 122 GISFNGNMDRVPPTQAIRAAOGLLACGVAOGLRSNYVLKGRDVRITSPGNOLYHL 181
 QY 183 QNMPHYR 189
 DB 182 QNMPHYR 188

RESULT 8
 AAY00770
 ID AAY00770 standard; Protein; 182 AA.
 XX
 XX AC AAY00770;
 XX
 XX DT 18-MAY-1999 (first entry)
 XX
 XX DE Mouse tag7 clone protein sequence.
 XX
 XX Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
 KW melanoma; leukaemia; apoptosis inducer; mouse.
 XX
 XX Mus sp.
 OS
 XX
 XX WO9902686-A1.
 XX
 XX 21-JAN-1999.
 XX
 XX 10-JUL-1998; 98WO-EP04287.
 XX
 XX 11-JUL-1997; 97US-0893764.
 XX
 XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 XX
 XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
 XX
 XX WPI; 1999-120867/10.
 XX N-PSDB; AAX21819.
 XX
 XX New nucleic acid encoding tag7 - used to inhibit tumour growth and
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
 PT leukaemia
 PT
 XX
 XX Claim 19; Fig 1; 138pp; English.
 XX
 XX This sequence is the murine tag7 of the invention. Cells containing
 CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
 CC to produce and purify antibodies; to inhibit growth of mammalian tumours,
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
 CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
 CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene
 CC mapping and detection of tag7 gene expression, and as primers. Antibodies
 CC against tag7 are used as reagents for detecting tag7, as an antagonist of
 CC tag7, for isolating tag7 and therapeutically to inhibit or delay tumour
 CC metastasis.
 CC
 XX Sequence 182 AA:
 SQ

Query Match 65.8%; Score 692; DB 20; Length 182;
 Best Local Similarity 67.2%; Pred. No. 6,9e-66;
 Matches 127; Conservative 21; Mismatches 33; Indels 8; Gaps 1;

QY 1 MLAMALPSLRIGAADETEPACCSPIVPRNEKALASECAQHLSLPRVYVVSHTAGS 60
 DB 1 MLFACALLALGLATLS-----CSFIVPRSEWRALPSECSRLGHVRVYVISHTAGS 52
 QY 61 SCNPASCOQOQARNVQYHMKTLGMCVGYNFLIGEDGLVYEGRGWNETGAHSGHLMNPM 120
 DB 53 FCNSPDCBQOQARNVQYHMKTLGMCVGYNFLIGEDGLVYEGRGWNETGAHSGHLMNPM 112
 QY 121 SIGISFNGNMDRVPPTQAIRAAOGLLACGVAOGLRSNYVLKGRDVRITSPGNOLYH 180

Matches 125; Conservative 20; Mismatches 35; Indels 9; Gaps 2

therapeutic value, and the identification of new secreted proteins is

CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 116 AA;

Query Match 47.08; Score 494; DB 21; Length 116;
Best Local Similarity 97.8%; Pred. No. 6.1e-45;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLAMALPSLRLGAAOETDPACCSPIVRNEMKALASCAQHLSPLRVYVSHRAGS 60
DB 6 MLAMALPSLRLGAAOETDPACCSPIVRNEMKALASCAQHLSPLRVYVSHRAGS 65

QY 61 SCNTPASCOOQARNVOHYHMKTLGMCVGYNFL 93
DB 66 SCNTXASCOOQARNVOHYHMKTLGMCVGYNFL 98

RESULT 11

ABB69757
ID ABB69757 standard; Protein; 184 AA.

AC ABB69757;

DE 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 36063.

XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

PN 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL13860.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

PS Disclosure: SEQ ID NO 36063; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (AB101840-ABL16175) and the encoded proteins
XX (AB101840-AB101840) and the encoded proteins
XX (AB101840-AB101840) and the encoded proteins
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 184 AA;

Query Match 39.4%; Score 414.5; DB 22; Length 184;
Best Local Similarity 47.2%; Pred. No. 3.8e-36;
Matches 77; Conservative 31; Mismatches 54; Indels 1; Gaps 1;

QY 28 IVPNEMKALASCAQHLSPLRVYVSHRAGSSCNTPASCOOQARNVOHYHMKTLGMD 87
DB 23 IISSEMGSRATSKTSANTISTAVIHHTAGNYSTAACTOIAOTDAYHMDISGMAD 82

QY 88 VGNFLLIGEDLVYEGRGMTGASGHLMPMSIGISFMGNVMDRVPPOAIRAOGGL 147
DB 83 IGVNPLIGDGNVYEGRGMTGASGHLMPMSIGISFMGNVMDRVPPOAIRAOGGL 141

QY 148 ACGVAOGLRNSNYVKGHRDVOFTLSPGNOLYHLIONPHYS 190
DB 142 SDAVSRCQIVSGYLLYGHROYGSTECPGTINWEIRTSNMKA 184

RESULT 12

ABB53271
ID ABB53271 standard; Protein; 241 AA.

AC ABB53271;

DE 12-FEB-2002 (first entry)

XX Human polypeptide #11.

XX Human; neotropic; neuroprotective; anticonvulsant; antidepressant;
XX neuroleptic; tranquiliser; antiarrhythmic; cardiac; antidiabetic;
XX antiinflammatory; antilipemic; hepatotropic; vitruicid; antidiabetic;
XX nephrotropic; anorectic; cytotatic; respiratory disease; liver disease;
XX renal disease; skeletal muscle disease; gastrointestinal disease;
XX placental disease; testicular cancer; male fertility; pancreatic disease.

OS Homo sapiens.

XX WO200181363-A1.

PN 01-NOV-2001.

PF 26-APR-2001; 2001WO-US13360.

PR 27-APR-2000; 2000US-199963P.

PR 11-MAY-2000; 2000US-203336P.

PR 25-MAY-2000; 2000US-207087P.

PR 26-MAY-2000; 2000US-207546P.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;

PI Lai Y, Xie Q;

XX WPI; 2002-041392/05.

DR N-PSDB; ABA90336.

PT Novel polypeptides and polynucleotides useful as a vaccine for

PT preventing and treating diseases associated the polypeptide, e.g.

PT Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,

XX asthma, amnesia

PS Claim 1; Page 79; 116pp; English.

XX The invention relates to an isolated polypeptide comprising a 277, 480,
XX 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
XX 844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,
XX 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
XX given in the specification. The polypeptides, modulators of the
XX polypeptides and antibodies against the polypeptides are useful for
XX treating diseases such as neurological and psychiatric diseases
XX including Alzheimer's, parasympathetic palsy, Huntington's disease,
XX myotonic dystrophy, anorexia and depression; cardiovascular diseases
XX including congestive heart failure, Hodgkin's disease and myocardial
XX infarction; respiratory diseases including asthma, chronic obstructive
XX pulmonary disease, cystic fibrosis and adult respiratory distress
XX syndrome; liver diseases including hypercholesterolaemia, cirrhosis,

CC that express zgpai, for screening expression libraries and as
 CC neutralizing antibodies or as antagonists to block zgpai activity
 CC in vitro and in vivo. Zgpai gene is also useful in gene therapy.
 XX

Sequence 369 AA;

Query Match 39.3%; Score 413; DB 22; Length 369;
 Best Local Similarity 40.0%; Pred. No. 1.4e-35;
 Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

QY 4 AWALPSLR----LGAQETEDPACCSPIVRNEMKALASCAQHLSPRLRYVVSHTAG 59
 DB 181 SYVPLLVKGENCLAPROKTSILKRAKCPGVPRSVWAGRETCPR-MTLPAKYGIITHAG 239
 QY 60 SSCNTPASCCQOQARNVQHYHKKTLGMCVGNFLLGEGDLYVEGRGMFTGASHGLNMP 119
 DB 240 RTCNIDSECRLLVLDIDQSFYIDRLKSCDIGNFLVGQGAITEGVMVQGS-STPGYDD 298
 QY 120 MSIGISFNGNMDRVPPOAIRAAQGLIACGVAAGALNSNYLKGHRDORTLSPGNOLY 179
 DB 299 IALGITFMGTGTGIPNNAALAAQDLIQCAMVKGYLTPNLLVGHSDVARTLSFGQALY 358
 QY 180 HLIONMPHYR 189
 DB 359 NIISTWPHFK 368

RESULT 14

ABB53272
 ID ABB53272 standard; Protein: 369 AA.

AC ABB53272;

DT 12-FEB-2002 (first entry)

DE Human polypeptide #12.

XX Human: neurotropic; neuroprotective; anticonvulsant; antidepressant;
 KW neuroleptic; tranquilizer; antiarrhythmic; cardiant; antiasthmatic;
 KW antiinflammatory; antidiabetic; hepatotropic; vitruicide; antidiabetic;
 KW nephrotropic; anorectic; cytostatic; vaccine; neurological disease;
 KW cardiovascular disease; respiratory disease; liver disease;
 KW renal disease; skeletal muscle disease; gastrointestinal disease;
 KW placental disease; testicular cancer; male fertility; pancreatic disease.

OS Homo sapiens.

PN MO200181363-A1.

PD 01-NOV-2001.

PF 26-APR-2001; 2001WO-US13360.

PR 27-APR-2000; 2000US-199963P.

PR 11-MAY-2000; 2000US-203336P.

PR 25-MAY-2000; 2000US-207087P.

PR 26-MAY-2000; 2000US-207546P.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;

PI Lai Y, Xie Q;

PI Novel polypeptides and polynucleotides useful as a vaccine for
 PT preventing and treating diseases associated the polypeptide, e.g.
 PT Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
 PT asthma, amnesias

PS Claim 1, Page 79-80; 116pp; English.

XX The invention relates to an isolated polypeptide comprising a 277, 480,
 CC 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
 CC 844, 782, 394, 447, 485, 286, 533, 490, 619, 490, 462, 255,
 CC 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
 CC given in the specification. The polypeptides, modulators of the
 CC polypeptides and antibodies against the polypeptides are useful for
 CC treating diseases such as neurological and psychiatric diseases
 CC including Alzheimer's, parasympathetic palsy, Huntington's disease,
 CC myotonic dystrophy, anorexia and depression; cardiovascular diseases
 CC including congestive heart failure, Hodgkin's disease and myocardial
 CC infarction; respiratory diseases including asthma, chronic obstructive
 CC pulmonary disease, cystic fibrosis and adult respiratory distress
 CC syndrome; liver diseases including hypercholesterolemia, cirrhosis,
 CC viral and nonviral hepatitis, Type II diabetes mellitus, and impaired
 CC glucose tolerance; renal disease including renal failure, acute tubular
 CC necrosis and glomerulonephritis; skeletal muscle diseases including
 CC Eulenburg's disease, hypoglycaemia and obesity; gastrointestinal
 CC diseases including myotonia congenita and intestinal obstruction; lymph
 CC diseases including lymphaglectasia; diseases of placenta including
 CC choriocarcinoma; diseases of testes including testicular cancer,
 CC male reproductive diseases including low testosterone and male
 CC infertility; and disease of pancreas including diabetic ketoacidosis,
 CC Type 1 and 2 diabetes and obesity. The present sequence is a
 CC polypeptide of the invention.

Sequence 369 AA;

Query Match 39.3%; Score 413; DB 23; Length 369;
 Best Local Similarity 40.0%; Pred. No. 1.4e-35;
 Matches 76; Conservative 42; Mismatches 66; Indels 6; Gaps 3;

QY 4 AWALPSLR----LGAQETEDPACCSPIVRNEMKALASCAQHLSPRLRYVVSHTAG 59
 DB 181 SYVPLLVKGENCLAPROKTSILKRAKCPGVPRSVWAGRETCPR-MTLPAKYGIITHAG 239
 QY 60 SSCNTPASCCQOQARNVQHYHKKTLGMCVGNFLLGEGDLYVEGRGMFTGASHGLNMP 119
 DB 240 RTCNIDSECRLLVLDIDQSFYIDRLKSCDIGNFLVGQGAITEGVMVQGS-STPGYDD 298
 QY 120 MSIGISFNGNMDRVPPOAIRAAQGLIACGVAAGALNSNYLKGHRDORTLSPGNOLY 179
 DB 299 IALGITFMGTGTGIPNNAALAAQDLIQCAMVKGYLTPNLLVGHSDVARTLSFGQALY 358
 QY 180 HLIONMPHYR 189
 DB 359 NIISTWPHFK 368

RESULT 15

AAE00692
 ID AAE00692 standard; Protein: 375 AA.

AC AAE00692;

DE 02-JUL-2001 (first entry)

DE Human full length granulocyte peptide homolog zgpai protein #1.

XX Human: granulocyte peptide A; GP-A; zgpai; cytostatic; antiinflammatory;
 KW vulnerary; dermatological; anti-microbial; gastrointestinal disease;
 KW pulmonary; dental caries; periodontal disease; gene therapy; AIDS;
 KW Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis;
 KW infection; urinary tract; respiratory; vaginal; lung; skin; cancer;
 KW ovarian; rectal; chromosome 1.

OS Homo sapiens.

OS Key Location/Qualifiers

FT Peptide 1..17 /label= Signal_peptide

FT Protein 18..375 /label= Mature_full_length_zgpai_protein_#1

FT	Region	/note- "this region functions as an immunogenic epitope
FT	Region	47..52
FT	Region	/label- Immunogenic-epitope
FT	Domain	/note- "this region is specifically claimed in claim 18"
FT	Region	58..189
FT	Region	/label- Domain_1
FT	Region	98..103
FT	Region	/label- Hydrophilic_region
FT	Region	123..128
FT	Region	/label- Hydrophilic_region
FT	Region	152..157
FT	Region	/label- Hydrophilic_region
FT	Region	156..161
FT	Region	/label- Hydrophilic_region
FT	Region	157..162
FT	Region	/label- Hydrophilic_region
FT	Region	158..163
FT	Region	/label- Hydrophilic_region
FT	Region	190..215
FT	Region	/label- Linker
FT	Domain	216..346
FT	Domain	/label- Domain_2
PN	WO200129224-A2.	
PN	26-APR-2001.	
PD	20-OCT-2000; 2000MO-US29177.	
PF	20-OCT-1999; 990S-0160712.	
PR	12-JUL-2000; 2000US-0218070.	
PR	(ZYMO) ZYMOGENETICS INC.	
PA	Conklin DC, Adler DA, Fox BA;	
PI	WPI; 2001-290918/30.	
DR	N-PSDB; AAD04004.	
XX	New granulocyte peptide homolog, zgap1 polypeptide, for research applications, diagnosis and treatment of cancer, periodontal, gastrointestinal disease, urinary tract, skin and lung infections	
XX	Claim 14; Page 105-106; 114pp; English.	
XX	The present sequence is human full length granulocyte peptide (GP-A) homolog, zgap1 protein. Zgap1 gene is located on human chromosome 1.	
CC	Zgap1 polypeptides are useful for producing antibodies which are useful for detecting cancer. Zgap1 polypeptides having anti-microbial activity are useful for treating dental caries, periodontal disease, thrush, detecting cancer. Zgap1 polypeptides having anti-microbial activity are useful for treating dental caries, periodontal disease, thrush, gastrointestinal disease, urinary tract infections, vaginal infections, respiratory infections, acquired immune deficiency syndrome (AIDS) and lung infections associated with cystic fibrosis and prevention of infection in skin and other epithelial wounds. Zgap1-cytokine fusion proteins are useful for enhancing in vivo killing of target tissues (epithelial cancers, and more specifically lung, ovarian and rectal cancers). Zgap1 polypeptides, fragments, fusion proteins or agonists are useful in in vitro studies of exogenous microorganism infections such as bacterial, viral or fungal infection and also to study epithelial cell defense induction in cell culture. Zgap1 antibodies, polynucleotides and polypeptides are useful for detection of zgap1 polypeptide, mRNA or anti-zgap1 antibodies, thus serving as markers for detecting genetic diseases or cancers. Zgap1 sequences are useful as diagnostics in forensic DNA profiling and as probes or primers to clone 5' non-coding regions of a zgap1 gene. Zgap1 antibodies are useful for tagging cells that express zgap1, for screening expression libraries and as neutralizing antibodies or as antagonists to block zgap1 activity in vitro and in vivo. Zgap1 gene is also useful in gene therapy.	
XX	Sequence 375 AA;	

[illegible]

Search completed: November 12, 2002, 02:22:14
Job time : 144 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 02:18:04 ; Search time 80 seconds
(without alignments)
229.521 Million cell updates/sec

Title: US-09-462-625-4

Sequence: 1 MLAWALPILRLGAQETE.....LSPGNLYHLIQNPHYRSP 191

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR-73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224	21.3	539	2	G70520
2	148.5	14.1	308	2	T35303
3	127	12.1	151	2	S07506
4	120	11.4	151	1	MOBPA7
5	91	8.7	879	2	F96558
6	82	7.8	268	2	T24371
7	79.5	7.6	267	2	H71878
8	79.5	7.6	267	2	B64636
9	78	7.4	432	2	C85087
10	77.5	7.4	277	2	I52825
11	77.5	7.4	282	2	S50031
12	77.5	7.4	345	2	T48111
13	77.5	7.4	356	2	D65096
14	77.5	7.4	356	2	A88124
15	77.5	7.4	356	2	H85968
16	76.5	7.3	379	2	AD3390
17	76.5	7.3	786	2	A47546
18	76	7.2	229	2	T03405
19	76	7.2	296	2	T35345
20	75.5	7.2	1467	2	T23950
21	75	7.1	764	2	S47569
22	75	7.1	803	2	A71475
23	74.5	7.1	313	2	JF0328
24	74.5	7.1	366	2	JF0328
25	74.5	7.1	313	2	JF0328
26	74.5	7.1	818	2	A83015
27	74.5	7.1	851	2	B75606
28	74	7.0	713	2	A35502
29	74	7.0	761	2	JC7821

30	74	7.0	1477	2	T18534	protein-tyrosine k
31	73.5	7.0	244	2	AC3533	l-fucose phospho
32	73.5	7.0	420	2	H75395	ABC transporter, p
33	73.5	7.0	439	2	A55520	YscW protein - Yer
34	73.5	7.0	439	2	T43585	Yops secretion pro
35	73.5	7.0	641	2	G85043	hypothetical prote
36	73.5	7.0	717	2	S31035	retrovirus-related
37	73.5	7.0	1159	2	T02866	hypothetical prote
38	73.5	7.0	1733	1	B45344	probable nuclear a
39	73.5	6.9	393	2	JF0180	phosphopentomutase
40	73	6.9	533	2	D75220	hypothetical prote
41	73	6.9	378	2	B55580	cardamomyl-phosphat
42	72.5	6.9	532	2	A72694	hypothetical prote
43	72.5	6.9	628	2	AC1469	transporter homolo
44	72.5	6.9	566	2	H83037	urease alpha subun
45	72	6.9				

ALIGNMENTS

RESULT 1
G70520
probable csp protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70520
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtrold, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; M01D:9829587; PMID:9634230
A:Accession: G70520
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-539 <COL>
A:Cross-references: GB:297188; GB:AL123456; M1D:93261805; PIDN:CAB10019.1; PID:el300C
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: csp

Query Match 21.3%; Score 224; DB 2; Length 539;
Best Local Similarity 29.0%; Pred. No. 1.3e-13;
Matches 64; Conservative 31; Mismatches 82; Indels 44; Gaps 8;

QY 5 WALPILRLGAQETEDPACCSPIVPRNEM---KALASECAOHLSPLRVVSHTAGSS 61
DB 183 WPPRAVYVIMAG---QPPA-----IISRAEMGADESILRCETPER-DKGVRAAVVHHTRGSN 233
QY 62 CMTPASCOOARNOVHYHMKTLGWCDCVGNFLIGEDGLVYEGRWNT---GAHSGHLW 117
DB 234 DYSPEASAGIVAYTYHSHKTLGWCIDVAYNALVDKYGQVFEASAGILTRVDEGFHTGG-F 292
QY 118 NMSGISFMGNVMDRVPPOAIRAOGILACGYA-----G 134
DB 293 NNTWGVAMIGNFDVAPPPIQIRIVGRLLGWLGMDDVDPRMSVDLQASGSSTYTPPGG 352
QY 155 ALRSNVYLKGRDVOFTLSPGNLYHLION---WPHYRSP 191
DB 353 AIALRLPAITFHTHDVGTDCPGNAAYAVMDEIRIDIAHFNDP 393

RESULT 2
T35303
hypothetical protein scsf7.1ac - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35303
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: 221574

A:Accession: T35303
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-308 <SEE>
 A:Cross-references: EMBL:AL096872; PIDN:CA51271.1; GSPDB:GN00070; SCOPDB:SC5F7.14C
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOPDB:SC5F7.14C

Query Match 14.1%; Score 148.5; DB 2; Length 308;
 Best Local Similarity 28.2%; Pred. No. 1.2e-06;
 Matches 66; Conservative 19; Mismatches 82; Indels 67; Gaps 12;

3 LAMAPSLRLGA-----AQETEDPACCSPTVPRNEMKALSEC 41
 66 LIGALPBLAVALFLCANGVERAADAEVPAATDRPAAPRDVPSVW---LQDA 122
 42 AQLSLPLRY-----VVVSHT--AGSSC-NTPASCOQANVQHYHMKTLGMCQVGYN 91
 123 ARAQAPAP-RYDDEVVAVFVHHTDPNGYDCAVPAI---LRGVYGGQGTGARMDMDIGYN 177
 92 FLIGEDSLYEGRGWNE---TGHSGLNMPMSIGISFGNMYDRVPTQAT----- 140
 178 FVYDRCGTYEGRAGGIDRPVTAHT-QGFNHRRTTGIALGTATAGVVPDELDAIAAV 236
 141 -----RAQGLACG---VAQALRSNVYLGKGRDVRQLSPGNOL 178
 237 AAMKLGSTGTDPRAKVALVSSNGLSRYAAGKTAMLPVAGHEDGYQTSCEPAL 290

RESULT 3
 S07506
 N:acetylmutamoyl-L-alanine amidase (EC 3.5.1.28) - phage T3
 N:Alternate names: gene 3.5 protein
 C:Species: phage T3
 C:Date: 30-Sep-1991 #sequence, revision 30-Sep-1991 #text_change 22-Jun-1999
 C:Accession: S07506
 R:Beck, P.J.; Gonzalez, S.; Ward, C.L.; Molineux, I.J.
 J. Mol. Biol. 210, 687-701, 1989
 A:Title: Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.
 A:Reference number: S07500; M01D:90133923; PMID:2614843
 A:Accession: S07506
 A:Molecule type: DNA
 A:Residues: 1-151 <BEC>
 A:Cross-references: EMBL:X17255; NID:g15682; PIDN:CA5133.1; PID:g15695
 C:Genetics:
 A:Gene: 3.5
 C:Superfamily: phage T7 N-acetylmutamoyl-L-alanine amidase
 C:Keywords: hydrolase

Query Match 12.1%; Score 127; DB 2; Length 151;
 Best Local Similarity 32.4%; Pred. No. 6.2e-05;
 Matches 36; Conservative 18; Mismatches 47; Indels 10; Gaps 4;

73 RNVOHYHMKTLGMCQVGYNFLIGEDGLYVEGRGWNFTGAHSHGLNMPMSIGISFGNMYD 132
 31 REIRQW-HKQGMNDVGYHFLIKRDGYEAGRDGLAVGSHAGK-YNNHSIGVCLVGGIDD 88
 133 RVP-----TPQAIRAAGCLACGVAAGALNSNVYLGKGRDVRQLSPGNOL 178
 89 KGRFDANFTPAQMOSLRSLVTLAK---YEGSVLRAHHVDVAPRACSFDL 136

RESULT 4
 M0BPAT
 N:acetylmutamoyl-L-alanine amidase (EC 3.5.1.28) - phage T7
 N:Alternate names: T7 lysoczyme
 C:Species: phage T7
 C:Date: 01-Sep-1981 #sequence, revision 01-Sep-1981 #text_change 18-Jun-1999
 C:Accession: G94615; D92866; S42302; S43502; A01001
 R:Dunn, J.J.; Thompson, K.
 submitted to the Nucleic Acid Sequence Database, September 1982
 A:Reference number: A94615

A:Accession: G94615
 A:Molecule type: DNA
 A:Residues: 1-151 <D01>
 R:Dunn, J.J.; Studier, F.W.
 J. Mol. Biol. 148, 303-330, 1981
 A:Title: Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the
 A:Reference number: A92866; M01D:82078034; PMID:7310871
 A:Accession: D92866
 A:Molecule type: DNA
 A:Residues: 1-151 <D02>
 R:Dunn, J.J.; Studier, F.W.
 J. Mol. Biol. 166, 477-535, 1983
 A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7
 A:Reference number: S42283; M01D:83241725; PMID:6864790
 A:Accession: S42302
 A:Molecule type: DNA
 A:Residues: 1-38, 'E', '40-151 <D04>
 A:Cross-references: EMBL:V01146
 R:Dunn, J.J.; Studier, F.W.
 submitted to the EMBL Data Library, October 1993
 A:Reference number: S43501
 A:Accession: S43502
 A:Molecule type: DNA
 A:Residues: 1-38, 'E', '40-118, 'V', '120-151 <D04>
 A:Cross-references: EMBL:V01146; NID:g431187; PIDN:CA24403.1; PID:g431190
 C:Comment: This late gene protein is not required for lysis but plays an important ro
 lly bound. The functional significance of this is not clear.
 C:Genetics:
 A:Gene: 3.5
 A:Map position: 26.77-27.90
 C:Superfamily: phage T7 N-acetylmutamoyl-L-alanine amidase
 C:Keywords: hydrolase

Query Match 11.4%; Score 120; DB 1; Length 151;
 Best Local Similarity 31.9%; Pred. No. 0.00029;
 Matches 36; Conservative 19; Mismatches 44; Indels 14; Gaps 5;

73 RNVOHYHMKTLGMCQVGYNFLIGEDGLYVEGRGWNFTGAHSHGLNMPMSIGISFGNMYD 132
 31 REIRQW-HKQGMNDVGYHFLIKRDGYEAGRDGLAVGSHAGK-YNNHSIGVCLVGGIDD 88
 133 RVP-----TPQAIRAAGCLACGVA--OGALRSNVYLGKGRDVRQLSPGNOL 178
 89 KGRFDANFTPAQMOSLRSLVTLAKYGA-----GLAHHVAVAKACPSFDL 136

RESULT 5
 F96558
 Probable protein kinase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence, revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F96558
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 Nansen, N.F.; Hughes, B.; Hultzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; M01D:21016719; PMID:11130712
 A:Accession: F96558
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-879 <STO>
 A:Cross-references: GB:AE005173; NID:99802791; PIDN:AA99960.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T14L22.12
 A:Map position: 1

Query Match 8.7%; Score 91; DB 2; Length 879;

Db 41 PPLPLGCLGETRDACGCCPMCARGEPEPCGGGGAGRGYCAPGMECVKSRKRKAGAGA 100

Matches 46; Conservative 17; Mismatches 76; Indels 77; Gaps 11;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 11, 2002, 23:30:29 ; Search time 307 Seconds

(without alignments)
5266.889 Million cell updates/sec

Title: US-09-462-625-3

Perfect score: 718
Sequence: 1 ctgagttactgagccagag.....ctctctccataaagatg 718

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

N.Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	100.0	718	20	AA21820
2	674	93.9	697	21	AA58104
3	674	93.9	697	21	AA37082
4	674	93.9	697	22	AA54356
5	673	93.7	726	21	AA80613
6	673	93.7	749	21	AA51719
7	349.4	48.7	688	18	AA78509
8	342.2	47.7	380	21	AA24249
9	316.6	44.1	549	20	AA21819

10	300.6	41.9	677	18	AA78510	Human granulocyte
11	231	32.2	285	21	AA80662	Human secreted pro
12	220	30.6	437	21	AA80663	Human secreted pro
13	147.4	20.5	558	23	ABL25883	Drosophila melanog
14	147.4	20.5	2558	23	ABL25882	Drosophila melanog
15	147.4	20.5	4320	23	ABL25876	Drosophila melanog
16	145.4	20.3	555	23	ABL25881	Drosophila melanog
17	145.4	20.3	2555	23	ABL25880	Drosophila melanog
18	145	20.2	639	23	ABL04835	Drosophila melanog
19	145	20.2	2639	23	ABL04834	Drosophila melanog
20	136.4	19.0	612	23	ABL26901	Drosophila melanog
21	132.2	18.4	1107	22	AAD04007	Human full length
22	132.2	18.4	1125	22	AAD04005	Human full length
23	117.6	16.4	726	24	ABA90336	Human polynucleoti
24	117.6	16.4	1110	22	AAD04006	Human full length
25	117.6	16.4	1110	24	ABA90337	Human polynucleoti
26	117.6	16.4	1128	22	AAD04004	Human full length
27	117.4	16.4	1876	21	AAAS1718	Wound healing tiss
28	112.6	15.7	4050	23	ABL25878	Drosophila melanog
29	98.2	13.7	1182	21	AAAS1717	Drosophila melanog
30	98.2	13.7	1191	21	AA265250	Keratinocyte pepti
31	97.8	13.6	522	19	AAV19147	Gene encoding pept
32	97.8	13.6	591	19	AAV19148	Gene encoding pept
33	97.8	13.6	753	19	AAV19149	Gene encoding the
34	94	13.1	1089	21	AAV15921	Human protein cion
35	94	13.1	1256	21	AAV15921	Human protein cion
36	94	13.1	1731	22	AA167206	Nucleotide sequenc
37	94	13.1	1905	22	AA167205	Nucleotide sequenc
38	92.4	12.9	1256	23	ABL14665	Drosophila melanog
39	90.2	12.6	1014	23	ABL07655	Drosophila melanog
40	87	12.1	561	23	ABL08767	Drosophila melanog
41	87	12.1	2561	23	ABL08766	Drosophila melanog
42	86.6	12.1	573	23	ABL15529	Drosophila melanog
43	86.6	12.1	2609	23	ABL15556	Drosophila melanog
44	86.6	12.1	2641	23	ABL15528	Drosophila melanog
45	85	11.8	4436	23	ABL07432	Drosophila melanog

ALIGNMENTS

RESULT 1	
AA21820	AA21820 standard; cDNA to mRNA; 718 BP.
ID	AA21820
AC	AA21820;
XX	
DT	18-MAY-1999 (first entry)
XX	
DE	Human tag7 clone coding sequence.
XX	
KW	Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
KW	melanoma; leukaemia; apoptosis inducer; human; ss.
XX	
OS	Homo sapiens.
XX	
PN	W09902686-A1.
XX	
PD	21-JAN-1999.
XX	
PF	10-JUL-1998; 98WO-EP04287.
XX	
PR	11-JUL-1997; 97US-0893764.
XX	
PA	(BOEH) BOEHRINGER INGELHEIM INT GMBH.
XX	
PI	Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E.
XX	
DR	WPI; 1999-120887/10.
XX	
DR	P-PSDB; AAY00771.
XX	
PT	New nucleic acid encoding tag7 - used to inhibit tumour growth and
PT	induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and

XX 02-DEC-1999; 99WO-US28551.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WT;
XX WPI; 2000-594320/56.
DR P-PSDB; AAB24022.
XX
XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression -
XX
XX Claim 50; Fig 3; 226pp; English.
XX
XX The present invention describes an antibody that binds to a human
CC Protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO3434; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO20262. (I) has
CC anticancer activity and can be used to diagnose tumors in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumors. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human PRO
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
XX SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
XX
XX Alignment Scores:
Pred. No.: 5.17e-100 Length: 697
Score: 1051.00 Matches: 191
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-462-625-4 (1-191) x AAC58104 (1-697)
QY 1 MetleuleuAlaTPAlaLeuProSerleuLeuArgleuGlyAlaAlaGlnGluThrGlu 20
DB 41 ATGCGCTGGCTGGGCTCCGCCAGCTCTCCATCGAGCGGCTGAGAGACAGAA 100
QY 21 ASPProAlaCysCysSerProIleValProArgAsnGluTyrIleValAlaLeuAlaSerGlu 40
DB 101 GACCGGCGCTGCTGACGCCCATAGTGGCCCGAAGAGAGTGAAGGCCCTGGCATAGAG 160
QY 41 CysAlaGlnHisLeuSerleuProleuArgTyrValAlaValSerHisThrAlaGlySer 60
DB 161 TGGCGCCAGCAGACCTGAGCTGCTTACGCTATGTGTGGTATCGACACAGCGGGCAGC 220
QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMet 80
DB 221 AGCTGCAACACCCCGCTGCTGCGACAGAGCGCCGGAATGTGACATACCAACATG 280
QY 81 LysThrLeuGlyTyrCysAspValGlyTyrAsnPhleuIleGlyGluAspGlyLeuVal 100
DB 281 AAGACACTGGGCTGTGCGACCTGGGCTACAACTCTCTATGTGGAGAAACAGCGCTCGTA 340
QY 101 TyrGlnGlyArgGlyTyrAsnPhThrGlyAlaHisSerGlyHisLeuThrAsnPromet 120

DB 341 TACGAGGCGCGGCTGGAACTTACAGGGTCCCACTCAGGTCACTATGAGAACCCCATG 400
QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
DB 401 TCCATTGGCATACGCTTCTGCGCAACTATCATGATCGGCTGCCCAACCCAGGCCATC 460
QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyr 160
DB 461 CGGCGAGCCCAAGGATGCTAGCTGCGGTGTGGCTCAGAGAGCCCTGAGGTCCAACTAT 520
QY 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
DB 521 GTGCTCAAGGACACCGGATGTGCAGCTACACTCTCTCAGGCAACACGCTCTACAC 580
QY 181 LeuIleGlnAsnTyrProHisTyrArgSerPro 191
DB 581 CTCATTCAGATTTGGCCACACTACCGCTCCCC 613
RESULT 2
AAA37082
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
XX 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (U0639) cDNA sequence SEQ ID NO:215.
XX
KM Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW Transmembrane; secretion; Immunoadhesion; pharmaceutical; screening;
XX ss.
XX Homo sapiens.
XX
PD WO200012708-A2.
XX
XX 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099815.
PR 15-SEP-1998; 98US-0099815.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.

DB 122 ATAGTCCCCGAGACAGTGGAGGCGCTGTGCATCAGAGTGGCGCCGACAGCTAGGCTG 181
QY 209 CCCTTACGCTATGTGTGTATGTCACACAGCGGGGAGCAGCTGCACACCCCGCTCG 268
DB 182 CCTTACGCTATGTGTGTATGTCACACAGCGGGGAGCAGCTGCACACCCCGCTCG 241
QY 269 TGGCAGCAGAGGCGCGGATGTGACGACCTACACATGAAACACAGGCGTGTGGAC 328
DB 242 TGGCAGCAGAGGCGCGGATGTGACGACCTACACATGAAACACAGGCGTGTGGAC 301
QY 329 GTGGGCTACACTTCCGATTGGAGAGAGGCGCTGTATACAGAGGCGCGTGTGGAC 388
DB 302 GTGGGCTACACTTCCGATTGGAGAGAGGCGCTGTATACAGAGGCGCGTGTGGAC 361
QY 389 TTACAGGCGTCCGACCTACGCTACCTATGAAACCCCATGTCATGGCATCAGCTCATG 448
DB 362 TTACAGGCGTCCGACCTACGCTACCTATGAAACCCCATGTCATGGCATCAGCTCATG 421
QY 449 GGCACACTACATGATGGGCGGTCGCCACACCCGACATCCGGGACGCCAGGCTCTACTG 508
DB 422 GGCACACTACATGATGGGCGGTCGCCACACCCGACATCCGGGACGCCAGGCTCTACTG 481
QY 509 GCCTGGGTGGTGGTCTGAGGAGCCCTGAGTCCACTATGTGTCAAAGAGACCGGGAT 568
DB 482 GCCTGGGTGGTGGTCTGAGGAGCCCTGAGTCCACTATGTGTCAAAGAGACCGGGAT 541
QY 569 GTGAGGCTACACTCTCTCCAGGACACAGCTACACCTCATCCAGATTTGGCCACAC 628
DB 542 GTGAGGCTACACTCTCTCCAGGACACAGCTACACCTCATCCAGATTTGGCCACAC 601
QY 629 TACCGCTCCCGCTGAGGCGCTGCTGATCCGACCCCATTCCTCCCTCCATGGCCAAA 688
DB 602 TACCGCTCCCGCTGAGGCGCTGCTGATCCGACCCCATTCCTCCCTCCATGGCCAAA 661
QY 689 ACCCAGCTGTCTCTCTCCCAATAAGATG 718
DB 662 ACCCAGCTGTCTCTCTCCCAATAAGATG 691

RESULT 3
AAA37082
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (UN0639) cDNA sequence SEQ ID NO:215.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 02-SEP-1998; 98US-0098936.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.

PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.

Db	602	TACGGCTCCCTGTAGGCCCTGCTGATGCCAGCCCATTCCTCCCTCCATGGCCAAA	661
QY	689	ACCCACATGTCTCTCTCTTCACATTAAGATG	718
Db	662	ACCCACATGTCTCTCTCTTCACATTAAGATG	691
RESULT 5			
AAA00613			
ID	AAA00613	standard; cDNA; 726 BP.	
XX	AAA00613;		
AC	21-NOV-2000	(first entry)	
XX			
DE	Human Htag7 secreted protein gene #8.		
XX			
KW	Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;		
KW	antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;		
KW	antitumor; vulnerrary; antiviral; antibacterial; antifungal;		
KW	immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;		
KW	multiple sclerosis; inflammatory disorder; inflammatory bowel disease;		
KW	Crohn's disease; nephritis; hyperproliferative disorder;		
KW	cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer		
KW	melanoma; lymphoma; wound healing; human; ss.		
XX			
OS	Homo sapiens.		
PN	MO200029435-A1.		
XX			
PD	25-MAY-2000.		
XX			
PE	27-OCT-1999; 99MO-US25031.		
PR	28-OCT-1998; 98US-0105971.		
XX			
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	N1 J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;		
PI	Greene JM;		
XX			
DR	WPI; 2000-387742/33.		
XX	P-PDB; AAB25583.		
XX			
PT	Isolated nucleic acid molecules encoding human secreted proteins are		
PT	used for the prevention, amelioration and treatment of autoimmune,		
PT	inflammatory, hyperproliferative and cardiovascular disorders, cancer,		
PT	wounds, and infectious diseases.		
XX			
XX	Claim 1; Figure 34; 803pp; English.		
XX			
CC	The present invention relates to 12 secreted human proteins and the		
CC	nucleotide sequences encoding them. The polynucleotide sequences given		
CC	in AAA00606-AA06623 encode the 12 secreted protein sequences given in		
CC	AAB25576-B25593. The human secreted proteins have various activities		
CC	dependent on the tissues in which they are expressed. Examples of the		
CC	activities of the proteins include: immunosuppressant;		
CC	anti-inflammatory; antiarthritic; antirheumatic; dermatological;		
CC	antiproliferative; antitumor; antirheumatic; dermatological;		
CC	antiviral; antibacterial; antiarteriosclerotic; antitumor; vulnerrary;		
CC	polymerides, agonists and antagonists may be used to treat prevent		
CC	and/or diagnose various disease, disorders and conditions examples of		
CC	which include: immune disorders e.g. Addison's disease, rheumatoid		
CC	arthritis, dermatitis, and multiple sclerosis; inflammatory disorders		
CC	e.g. inflammatory bowel disease, Crohn's disease and nephritis;		
CC	hyperproliferative disorders such as paraproteinemias and purpura;		
CC	cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis		
CC	cancer e.g. melanoma and lymphoma. The proteins and polynucleotide		
CC	sequences may also be used in wound healing and the treatment of		
CC	infectious diseases. The human secreted protein gene #8 and protein		
CC	sequences are represented in sequences AAA00613 and AAB25583. Sequences		
CC	AAA00662-AA0663 represent genes related to the secreted protein gene#8.		
XX			

PT Leukemia

XX Claim 11; Page 126-127; 138pp; English.

CC This sequence encodes the human tag7 of the invention. Cells containing
 CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
 CC to produce and purify antibodies; to inhibit growth of mammalian tumors,
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
 CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
 CC weight marker. The tag7 polypeptide inhibits tumor growth and induces
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene
 CC mapping and detection of tag7 gene expression, and as primers. Antibodies
 CC against tag7 are used as reagents for detecting tag7; as an antagonist of
 CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumor
 CC metastasis.

XX Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;

Query Match 100.0%; Score 718; DB 20; Length 718;

Best Local Similarity 100.0%; Pred. No. 1.4e-172;

Matches 718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGTTACTGGGCGCCAGAGGCTGGGCCCTGGACATGTACTGACAGCCACTATGTCGG 60
 DB 1 CTGAGTTACTGGGCGCCAGAGGCTGGGCCCTGGACATGTACTGACAGCCACTATGTCGG 60
 QY 61 CCAGCTATGCTGCTTGGCCGCTGCTCCGACGCTCTCCGACGCTCTGACGAGCGGCTCAGGA 120
 DB 61 CCAGCTATGCTGCTTGGCCGCTGCTCCGACGCTCTCCGACGCTCTGACGAGCGGCTCAGGA 120
 QY 121 GACAGAGAGCCGCGCTGCTGACAGCCCATATGTCCTCCGAGAGAGTGAAGGCGCTGGC 180
 DB 121 GACAGAGAGCCGCGCTGCTGACAGCCCATATGTCCTCCGAGAGAGTGAAGGCGCTGGC 180
 QY 181 ATGAGTGGGCGCCAGACACTGAGCCTGACCTTACGCTATGTTGGTGTATGACACAGCG 240
 DB 181 ATGAGTGGGCGCCAGACACTGAGCCTGACCTTACGCTATGTTGGTGTATGACACAGCG 240
 QY 241 GGGCAGCAGCTGCAACACCCCGCTGCTGCGACAGCAGAGCCCGGAAATGAGCACTA 300
 DB 241 GGGCAGCAGCTGCAACACCCCGCTGCTGCGACAGCAGAGCCCGGAAATGAGCACTA 300
 QY 301 CCACATGAAGACACTGGGCTGCTGCGACAGTGGGCTACAACTTCTGATTGGAGAGACGG 360
 DB 301 CCACATGAAGACACTGGGCTGCTGCGACAGTGGGCTACAACTTCTGATTGGAGAGACGG 360
 QY 361 GGTGATGAGAGGGGCGGCTGCGAAGTCAAGGGTCCCACTCAGGTCACTTATGAA 420
 DB 361 GGTGATGAGAGGGGCGGCTGCGAAGTCAAGGGTCCCACTCAGGTCACTTATGAA 420
 QY 421 CCCCATGCTCATTGGCATGAGTTCATGGGCACTACATGATGGGTGCCACACCCCA 480
 DB 421 CCCCATGCTCATTGGCATGAGTTCATGGGCACTACATGATGGGTGCCACACCCCA 480
 QY 481 GGGCATCCGGGAGCCAGGCTTACTGGCTGCGGTGGTGGCTCAGGAGACCCCTGAGGTC 540
 DB 481 GGGCATCCGGGAGCCAGGCTTACTGGCTGCGGTGGTGGCTCAGGAGACCCCTGAGGTC 540
 QY 541 CAACATATGCTCAAGAGACCGGGATGTCAGGCTACACTCTCCAGGCAACCAAGCT 600
 DB 541 CAACATATGCTCAAGAGACCGGGATGTCAGGCTACACTCTCCAGGCAACCAAGCT 600
 QY 601 CTACACATCTATCAGAAATGGGCACTACAGGCTCCCTGAGGCGCTGATCGCA 660
 DB 601 CTACACATCTATCAGAAATGGGCACTACAGGCTCCCTGAGGCGCTGATCGCA 660
 QY 661 CCCCATTTCTCCCTCCCATGGGCAAAAACCCCACTGTCTCTTCTCAATAAAGATG 718
 DB 661 CCCCATTTCTCCCTCCCATGGGCAAAAACCCCACTGTCTCTTCTCAATAAAGATG 718

RESULT 2

AAC58104 ID AAC58104 standard; cDNA: 697 BP.

XX AAC58104;

DT 25-JAN-2001 (first entry)

DE Human PRO1269 nucleotide sequence SPO ID NO:6.

KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
 KM identification; tumourigenesis; anticancer; detection; ss.

OS Homo sapiens.

PN W0200053750-A1.

PD 14-SEP-2000.

PE 02-DEC-1999; 99WO-US28551.

PR 08-MAR-1999; 99WO-US05028.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 01-DEC-1999; 99WO-US28634.

PA (GENT) GENENTECH INC.

PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WL;

DR WPI: 2000-594320/56.

DR P-PSDB; AAB24022.

PT Antibodies specific for PRO polypeptides, used to diagnose and inhibit

PT the growth of tumors in mammals, and to identify inhibitors of PRO
 PT polypeptide activity or expression

XX Claim 50; Fig 3; 226pp; English.

XX The present invention describes an antibody that binds to a human
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
 CC PRO3434; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
 CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
 CC anticancer activity and can be used to diagnose tumors in mammals, by
 CC detecting complex formation when the antibody is contacted with test
 CC cells. Increased expression of genes encoding (I) can also be detected
 CC to diagnose tumors. Agents which inhibit the activity of (I),
 CC especially the antibodies, or an antisense oligonucleotide which
 CC hybridises to genes encoding (I), can be used to inhibit tumor growth,
 CC preferably by inducing cell death. Methods from the present invention
 CC can be used to identify compounds which inhibit the biological activity
 CC of (I). AAC58103 to AAC58102 represent PCR primers and hybridisation
 CC probes used in examples from the present invention for human PRO
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
 CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.

XX Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Query Match 93.9%; Score 674; DB 21; Length 697;

Best Local Similarity 98.6%; Pred. No. 2e-161;

Matches 680; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 29 CCTGGACATGTACTGAGCAGCACTATGTCCTCCGCTATGCTGCTGGGCTCTC 88
 DB 2 CCGGACCTCTGCGCCCTGCACTATGTCCTCCGCTCTATGCTGCTGGGCTCTC 61
 QY 89 CCCAGCTCTCTTGCAGTGGAGGCGCTCAGAGAGAGAACCCGGGCTCTGAGGCC 148
 DB 62 CCGAGCTCTCTTGCAGTGGAGGCGCTCAGAGAGAGAACCCGGGCTCTGAGGCC 121
 QY 149 ATATGTCCTCCGAGAGAGTGAAGGCCCTGGCATCAGATGCGCCAGCACTGAGCCTG 208

PR	10-SEP-1998;	98US-0099741
PR	10-SEP-1998;	98US-0099754
PR	10-SEP-1998;	98US-0099763
PR	10-SEP-1998;	98US-0099792
PR	10-SEP-1998;	98US-0099808
PR	10-SEP-1998;	98US-0099812
PR	10-SEP-1998;	98US-0099815
PR	10-SEP-1998;	98US-0099816
PR	15-SEP-1998;	98US-0100385
PR	15-SEP-1998;	98US-0100388
PR	15-SEP-1998;	98US-0100390
PR	16-SEP-1998;	98US-0100584
PR	16-SEP-1998;	98US-0100627
PR	16-SEP-1998;	98US-0100661
PR	16-SEP-1998;	98US-0100662
PR	16-SEP-1998;	98US-0100664
PR	17-SEP-1998;	98US-0100683
PR	17-SEP-1998;	98US-0100684
PR	17-SEP-1998;	98US-0100710
PR	17-SEP-1998;	98US-0100711
PR	17-SEP-1998;	98US-0100919
PR	17-SEP-1998;	98US-0100930
PR	18-SEP-1998;	98US-0100939
PR	18-SEP-1998;	98US-0100849
PR	18-SEP-1998;	98US-0101014
PR	18-SEP-1998;	98US-0101068
PR	22-SEP-1998;	98US-0101071
PR	23-SEP-1998;	98US-0101471
PR	23-SEP-1998;	98US-0101472
PR	23-SEP-1998;	98US-0101474
PR	23-SEP-1998;	98US-0101475
PR	23-SEP-1998;	98US-0101476
PR	23-SEP-1998;	98US-0101477
PR	23-SEP-1998;	98US-0101479
PR	24-SEP-1998;	98US-0101738
PR	24-SEP-1998;	98US-0101741
PR	24-SEP-1998;	98US-0101743
PR	24-SEP-1998;	98US-0101915
PR	24-SEP-1998;	98US-0101916
PR	29-SEP-1998;	98US-0102207
PR	29-SEP-1998;	98US-0102307
PR	29-SEP-1998;	98US-0102331
PR	30-SEP-1998;	98US-0102484
PR	30-SEP-1998;	98US-0102571
PR	30-SEP-1998;	98US-0102571
PR	01-OCT-1998;	98US-0102684
PR	01-OCT-1998;	98US-0102687
PR	02-OCT-1998;	98US-0102965
PR	06-OCT-1998;	98US-0103258
PR	06-OCT-1998;	98US-0103449
PR	07-OCT-1998;	98US-0103314
PR	07-OCT-1998;	98US-0103315
PR	07-OCT-1998;	98US-0103328
PR	07-OCT-1998;	98US-0103395
PR	07-OCT-1998;	98US-0103396
PR	08-OCT-1998;	98US-0103401
PR	08-OCT-1998;	98US-0103433
PR	08-OCT-1998;	98US-0103678
PR	08-OCT-1998;	98US-0103679
PR	14-OCT-1998;	98US-0103711
PR	20-OCT-1998;	98US-0104257
PR	20-OCT-1998;	98US-0104987
PR	20-OCT-1998;	98US-0105000
PR	21-OCT-1998;	98US-0105002
PR	22-OCT-1998;	98US-0105104
PR	22-OCT-1998;	98US-0105169
PR	26-OCT-1998;	98US-0105266
PR	26-OCT-1998;	98US-0105693
PR	26-OCT-1998;	98US-0105694

PR 27-OCT-1998; 98US-0105807.
 PR 27-OCT-1998; 98US-0105881.
 PR 27-OCT-1998; 98US-0105882.
 PR 27-OCT-1998; 98US-0106062.
 PR 28-OCT-1998; 98US-0106023.
 PR 28-OCT-1998; 98US-0106029.
 PR 28-OCT-1998; 98US-0106030.
 PR 28-OCT-1998; 98US-0106033.
 PR 28-OCT-1998; 98US-0106178.
 PR 29-OCT-1998; 98US-0106248.
 PR 29-OCT-1998; 98US-0106384.
 PR 29-OCT-1998; 98US-0106500.
 PR 30-OCT-1998; 98US-0106464.
 PR 03-NOV-1998; 98US-0106556.
 PR 03-NOV-1998; 98US-0106902.
 PR 03-NOV-1998; 98US-0106905.
 PR 03-NOV-1998; 98US-0106919.
 PR 03-NOV-1998; 98US-0106934.
 PR 10-NOV-1998; 98US-0107783.
 PR 17-NOV-1998; 98US-0108775.
 PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108788.
 PR 17-NOV-1998; 98US-0108801.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108925.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.
 PR 18-NOV-1998; 98US-0108904.
 PR (GETH) GENENTECH INC.
 PR Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
 PR WPI: 2000-237871/20.
 PR P-PSDB: AAY99400.
 PR New mammalian DNA sequences encoding transmembrane, receptor or
 PR secreted PRO polypeptides, useful for screening of potential peptide or
 PR small molecule inhibitors of the relevant receptor/ligand interactions
 PR
 PS Claim 2; Fig 121; 773pp; English.
 XX
 CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
 CC receptor or secreted PRO polypeptides given in AAY9340 to AAY99462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding them have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX
 SO Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Query Match 93.9%; Score 674; DB 21; Length 697;
 Best Local Similarity 98.6%; Pred. No. 2e-161;
 Matches 680; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 29 CCGGACATGTAACCTGACCACTATGTCCTCCGCGCTATGCTGCTGGGCTCTC 88
 DB 2 CCGGACCTCTTGAACCTGACCACTATGTCCTCCGCGCTATGCTGCTGGGCTCTC 61
 QY 89 CCGGACCTCTTGAACCTGACCACTATGTCCTCCGCGCTATGCTGCTGGGCTCTC 148

DB 62 CCGGACCTCTTGAACCTGACCACTATGTCCTCCGCGCTATGCTGCTGGGCTCTC 121
 QY 149 ATAGTCCCGGAAAGAGTGAAGCCCTGGCATCAGAGTGGCCGACCTGAGCCCTG 208
 DB 122 ATAGTCCCGGAAAGAGTGAAGCCCTGGCATCAGAGTGGCCGACCTGAGCCCTG 181
 QY 209 CCCTTACGCTATGTGTGTGTATGTCACACAGGCGGGGACAGCTGCACACCCCGCTCG 268
 DB 182 CCCTTACGCTATGTGTGTGTATGTCACACAGGCGGGGACAGCTGCACACCCCGCTCG 241
 QY 269 TGGCAGAGAGAGCGCGCGGAAATGTCACACATCAATGAAGACATCTGGCTGTCGAC 328
 DB 242 TGGCAGAGAGAGCGCGCGGAAATGTCACACATCAATGAAGACATCTGGCTGTCGAC 301
 QY 329 GTGGGCTACAACTTCCTGATTTGGAAGAGCGGCTGATACGAGGCGCTGGCTGAAAC 388
 DB 302 GTGGGCTACAACTTCCTGATTTGGAAGAGCGGCTGATACGAGGCGCTGGCTGAAAC 361
 QY 389 TTCAAGGGGTGCCCACTGAGTCACTTATGGAACCCCATGTCATGTCATGAGCTTCATG 448
 DB 362 TTCAAGGGGTGCCCACTGAGTCACTTATGGAACCCCATGTCATGTCATGAGCTTCATG 421
 QY 449 GGCAACTACATGATGGGTGGTGGCCACACCCGAGCCCATCGGAGCGAGGCTTACTG 508
 DB 422 GGCAACTACATGATGGGTGGTGGCCACACCCGAGCCCATCGGAGCGAGGCTTACTG 481
 QY 509 GCGTGCAGGTGTGCTGACGAGGAGCCCTGAGTCACTATGTCCTCAAGACACCGGAT 568
 DB 482 GCGTGCAGGTGTGCTGACGAGGAGCCCTGAGTCACTATGTCCTCAAGACACCGGAT 541
 QY 569 GTGACAGGTACACTCTTCCAGGACACAGCTTACACCTCATCAGATTTGCCACAC 628
 DB 542 GTGACAGGTACACTCTTCCAGGACACAGCTTACACCTCATCAGATTTGCCACAC 601
 QY 629 TACGCTCCCGCTGAGGAGCCCTGCTGATCCGACCCCATTCCTCCCTCCATGGCCAAA 688
 DB 602 TACGCTCCCGCTGAGGAGCCCTGCTGATCCGACCCCATTCCTCCCTCCATGGCCAAA 661
 QY 689 ACCCCACTGTCTCTTCCATTAAGATG 718
 DB 662 ACCCCACTGTCTCTTCCATTAAGATG 691
 RESULT 4
 ID AAF54356 standard; DNA; 697 BP.
 AC AAF54356.
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE DNA encoding protein of the invention #61.
 XX
 KW Secreted; transmembrane; gene therapy; ss.
 OS Unidentified.
 OS
 PN WO200078961-A1.
 PD 28-DEC-2000.
 PF 18-FEB-2000; 2000MO-US04342.
 PR 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99MO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99MO-US28313.
 PR 02-DEC-1999; 99MO-US28551.
 PR 16-DEC-1999; 99MO-US30095.
 PR 05-JAN-2000; 2000MO-US00219.

ID	AAZ42549 standard; cDNA: 380 BP.
XX	AAZ42549;
AC	AAZ42549;
DT	01-FEB-2000 (first entry)
XX	
DE	Human 5' EST isolated from a cDNA library SEQ ID NO:308.
XX	
KM	Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;
KM	gene therapy; chromosome mapping; upstream regulatory sequence;
KM	forensic; location; development; protein synthesis; stability;
KM	regulation; identification; ss.
OS	Homo sapiens.
XX	
PN	W09953051-A2.
XX	
PD	21-OCT-1999.
XX	
PF	09-APR-1999; 99WO-IB00712.
XX	
XX	09-APR-1998; 98US-0057719.
PR	28-APR-1998; 98US-0069047.
XX	
PA	(GEST) GENSET.
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	
PI	WPI: 2000-038446/03.
DR	P-PSDB; AAY64935.
XX	
PT	Novel secreted protein 5' expressed sequence tag sequences used in
XX	diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX	
PS	Claim 1; Page 336; 837pp; English.
CC	AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC	sequences, corresponding to human secreted proteins. AAY4651 to
CC	AAY6538 represent the EST-related proteins corresponding to AAZ42265 to
CC	AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC	products. They can be used to identify and isolate 5' untranslated
CC	regions (UTRs) and upstream regulatory regions which control the
CC	location, development stage, rate, and quantity of protein synthesis, as
CC	well as stability of mRNA. The ESTs are also useful as probes for
CC	chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC	also be used in forensic procedures to identify individuals, or in
CC	diagnostic procedures to identify individuals having genetic diseases
CC	resulting from abnormal gene expression. The products may also be used in
CC	gene therapy protocols. The nucleic acids encoding signal peptides can be
CC	used for directing extracellular secretion of a polypeptide or the
CC	insertion of a polypeptide into a membrane, or importing a polypeptide
CC	into a cell. The proteins encoded by the EST sequences may be useful in
CC	treating a variety of human conditions. Secreted proteins have
CC	therapeutic value, and the identification of new secreted proteins is
CC	valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC	sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 380 BP; 67 A; 130 C; 114 G; 63 T; 6 other:
07	Query Match 47.7%; Score 342.2; DB 21; Length 380;
Db	Best Local Similarity 95.0%; Pred. No. 2.6e-77;
144	Matches 358; Conservative 5; Mismatches 13; Indels 1; Gaps 1
07	24 GGGGCCCTGGAGATGTAAGCTGACGACGACATATGCCCGGCGCTATAGCGCTTGGG 83
Db	4 GGTTCGCCGGCCCTGCGCCGCTGACATATGTCGCCGCTATAGCTGTGCTGGG 63
07	84 CTCCTCCCAAGCCTCTTCGACTCGAGCGGGCTCAGAGAGACAGAGACCCTGCTGTGA 143
Db	64 CTCCTCCCAAGCCTCTTCGACTCGAGCGGGCTCAGAGAGACAGAGACCCTGCTGTGA 123
07	144 GCCCATATGTCGCCCGGAGAGAGTGGAGAGCCCTGGCATCAAGTGCGCCACACCTGA 203

Db	124	GCCCATATGTCGCCCGGAAGAGTGGAAAGCCCTGGCATCAGATGCGCCCGACACTGA	183
QY	204	GCCTGCGCTTACGCTATGTGTGGTATGGCACACAGCGCGGCGACGACGCTCAACACCCCGG	263
Db	184	GCTTCGCCCTTACGCTATGTGTGGTATGGCACACGCGCGGCGACGCTCAACACCCCGG	243
QY	264	CCTGTGTCACAGCAGCGCCGGAA\TGTCAGACACTTACACATGAAGACACTGGGCTGCT	323
Db	244	CTGTGTGCCACAGCAGCGCCGGAA\TGTCAGACACTTACACATGAAGACACTGGGCTGCT	303
QY	324	GCAGCTGGGGCTTACAACTTCTCT-GATTGGAGAAGACGGGCTGCTATACGAGGCGCGTGC	382
Db	304	GCGAGCTGGGGCTTACAACTTCTCTNGATTGGAGAAGACGGGCTGCTATACAGGGCCGTGGA	363
QY	383	TGGAACCTTCACGCGGTGC 399	
Db	364	TGGAACCTTCACGCGGTGC 380	
RESULT 9			
AXX21819			
ID	AXX21819	standard; cDNA; 549 BP.	
XX	AXX21819;		
AC			
XX	18-MAY-1999	(first entry)	
DT			
XX			
DE		Mouse tag7 clone coding sequence.	
XX			
KM		Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;	
XX		melanoma; leukaemia; apoptosis inducer; mouse; ds.	
OS		Mus sp.	
XX			
PN		WO9902686-A1.	
PD		21-JAN-1999.	
XX			
PF		10-JUL-1998; 98WO-EP04287.	
XX			
PR		11-JUL-1997; 97US-0893764.	
XX			
PA		(BOEH) BOEHRINGER INGELHEIM INT GMBH.	
XX			
PI		Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;	
XX			
DR		WPI; 1999-120887/10.	
DR		P-PSDB; AAY00770.	
XX			
PT		New nucleic acid encoding tag7 - used to inhibit tumour growth and	
XX		induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and	
XX		leukaemia	
PS		Claim 3; Fig 1; 138pp; English.	
CC			
XX		This sequence encodes the murine tag7 of the invention. Cells containing	
CC		the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used	
CC		to produce and purify antibodies; to inhibit growth of mammalian tumours,	
CC		especially for treating carcinoma (of liver, ovary, breast, cervix, lung,	
CC		prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head	
CC		and neck, squamous cell carcinoma or leiocarcinoma), sarcoma (Kaposi's,	
CC		osteio- or fibro-sarcomas), melanoma or leukaemia; and as a molecular	
CC		weight marker. The tag7 polypeptide inhibits tumour growth and induces	
CC		apoptosis. The tag7 coding sequences are also useful as probes for gene	
CC		mapping and detection of tag7 gene expression, and as primers. Antibodies	
CC		against tag7 are used as reagents for detecting tag7; as an antagonist of	
CC		tag7; for isolating tag7 and therapeutically to inhibit or delay tumour	
XX		metastasis.	
SO		Sequence 549 BP; 117 A; 161 C; 151 G; 120 T; 0 other;	
Query Match	44.1%;	Score 316.6; DB 20; Length 549;	
Best Local Similarity	77.2%;	Pred No. 8; 8e-71;	

Query Match	32.2%	Score 231	DB 21	Length 285
Best Local Similarity	99.2%	Pred. NO. 3.9e-49		
Matches 242	Conservative	0	Mismatches 1	Indels 1
			Gaps 1	

475 ACCCGAGCGCATCCGGCGACGCCAGGGTCTACTGCGCTCGCGGTGTGGCTCAGGAGCCCT 534

|||||

Db	1	ACCCAGGGCCAACTCCGGGNAAGCCCAAGGCTCTACTGCGCTTCGTGGCTCAGGGAGC- CCT	59
Qy	533	GAGGTCCAACTATGTGCTCCAAAGGACACCGGAGTGTGACGCTACACTCTCTCAGGCCAA	594
Db	60	GAGGTCCAACTATGTGCTCCAAAGGACACCGGAGTGTGACGCTACACTCTCTCAGGCCAA	119
Qy	595	CCAGCTTCACCAACCGTCATCCAGAAATTTGGCCGACACTACCGCTCCCGCCGAGAGCCCTGCTGA	654
Db	120	CCAGCTTCACCAACCGTCATCCAGAAATTTGGCCGACACTACCGCTCCCGCCGAGAGCCCTGCTGA	179
Qy	655	TCGCGACCCCAATTCCTCCCTCCCATGGCCAAAAACCCCACTGTCTCTCTCCAAATAAA	714
Db	180	TCGCGACCCCAATTCCTCCCTCCCATGGCCAAAAACCCCACTGTCTCTCTCCAAATAAA	239
Qy	715	GATG 718	
Db	240	GATG 243	
RESULT 12			
AAA80663			
ID	AAA80663 standard; cDNM: 457 BP.		
XX	AAA80663;		
XX	21-NOV-2000 (first entry)		
DE	Human secreted protein gene #8 related gene HCDP40R SEQ ID #116.		
KM	Secreted protein: immunosuppressant; anti-inflammatory; antiarthritic;		
KM	antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;		
KM	anticancer; vulnerary; antiviral; antibacterial; antifungal;		
KM	immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;		
KM	multiple sclerosis; inflammatory disorder; inflammatory bowel disease;		
KM	Crouh's disease; nephritis; hyperproliferative disorder;		
KM	cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;		
KM	melanoma; lymphoma; wound healing; human; ss.		
XX	Homo sapiens.		
OS	MO200029435-AL.		
XX	25-MAY-2000.		
XX	27-OCT-1999; 99WO-DS25031.		
PF	28-OCT-1998; 98US-0105971.		
PR	(HUMA-) HUMAN GENOME SCI INC.		
XX	Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;		
XX	Greene JM;		
DR	WPI: 2000-387742/33.		
XX	Isolated nucleic acid molecules encoding human secreted proteins are		
PT	used for the prevention, amelioration and treatment of autoimmune,		
PT	inflammatory, hyperproliferative and cardiovascular disorders, cancer,		
PT	wounds, and infectious diseases -		
PS	Disclosure: Page 759-760; 803pp; English.		
CC	The present invention relates to 12 secreted human proteins and the		
CC	nucleotide sequences encoding them. The polynucleotide sequences given		
CC	in AAA80606-880623 encode the 12 secreted protein sequences given in		
CC	AA825576-825593. The human secreted proteins have various activities		
CC	dependent on the tissues in which they are expressed. Examples of the		
CC	activities of the proteins include: immunosuppressant;		
CC	anti-inflammatory; antiarthritic; antirheumatic; dermatological;		
CC	antiproliferative; antiarteriosclerotic; anticancer; vulnerary;		
CC	antiviral; antibacterial; and antifungal activity. The proteins,		
CC	polypeptides, agonists and antagonists may be used to treat prevent		
CC	and/or diagnose various disease, disorders and conditions examples of		

Query Match	30.6%	Score 220;	DB 21;	Length 457;
Best Local Similarity	84.5%	Pred. NO. 2.7e+46;		
Matches 283; Conservative	0;	Mismatches 47;	Indels 5;	Gaps 4

RESULT 13
ABL25883
ID ABL25883 standard; DNA; 558 BP.

"C" ISOLATED NUCLEIC ACID DETECTION REAGENT FOR DETECTING 1000 OR MORE

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161176-AB130511), expressed DNA
CC sequences (AB018400-AB161175) and the encoded proteins
CC (AAB57737-AB872012).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

50 Sequence 558 BP; 108 A; 199 C; 159 G; 92 T; 0 other;

Query Match	20.5%;	Score 147.4;	DB 23;	Length 558;
Best Local Similarity	57.3%;	Pred. No. 7.4e-28;		
Matches 287;	Conservative	0;	Mismatches 211;	Indels 3;
				Gaps 1

QY	129	ACCGGGCTGCTGAGCCCATATATGCCCCGGAACGAGTGGAGAGCCCTGGCATCAGAGT	188
Db	50	ACATGGCCACAGGGGGTATATGTCTCCACAGCGGAGTGGGGGTGTCCGGGCGCCAAAT	109
QY	189	GGCGCCAGCAACCTCAGGCTGGCCCTTACGTATGtGGGGATATGCCAACAACGGCGGCGACA	248
Db	110	GGACCGTAGGCTGTGGGCACTACCTCAGCTACGCCATCATCCACACACGCGGGCTCT	169
QY	249	GCTGGCAACACCCCGCCTCTGGCAGCAGCAGGCGCCGGAATGTGCACACTCAACACATGA	308
Db	170	ACTCGAGAGACCCGtGGGCCATGGACACGCCGCTGCTGCAGAGCGTCCAGAACTACCACTGG	229
QY	309	AGACACTGGGCTGTGTGGCAGCTGGGGCTACAACTTCTCTGAATTGGAGAGACGGGCTGTAT	368
Db	230	ACTTCCTCGGGCTGGCCCGACATCATGGCTACAACTTCTGATCGGGGGAGAGGCAAGCTGT	289
QY	369	ACGAGGGGCGTGGCTGGAACTTACAGGGGTCCCACTCAGGCTACTTATGGAAACCCATGT	428
Db	290	ACGAGGGGCGTGGCTGGAAACATATGGGCGCCACACCGC---CCGAGTGGAAACCCCTACA	346
QY	429	CCATTGGCATCAGCTTCATGGGCAACATCTGATGGGGTGGCCACACCCACAGCCATCC	488
Db	347	GCACTGGCATCAGCTTCTCTGGCAACTACACATGGGACACCTCTGGAGCCCAACATGATCT	406
QY	489	GGGCAAGCCAGGGTCTACTGTGGCCCTGGGAGTGTGGCTCAGGAGCGCTGAGGTCCAACTATG	548
Db	407	CCGCGCGCCAGCAGCTGTCTCAAGCAGCGCGCTCAACCTGTGGCCAGCTCAGCTCCGGCTACA	466
QY	549	TGCTCAAAAGACACCGGGATGTGCACGTCACACTCTCTCCAGGCAACCAAGCTTACACAC	608
Db	467	TCTGTACGCTCATCCGCAAGTCAAGCGCACCGAATGCCCCCGCACCCACATCTGGAAAG	526
QY	609	TCATCCAGATTTGGCCACAT 629	
Db	527	AGATCCGCGGGCTGTCCCACT 547	

RESULT 14
ABL25882/C
ID ABL25882 standard; DNA; 2558 BP.

26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 29119.

DE

XX Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX

```

OS Drosophila melanogaster.
XX
XX PN W0200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX DR WPI; 2001-656860/75.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX
XX PS Claim 1; SEQ ID NO 29119; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2558 BP; 596 A; 588 C; 644 G; 730 T; 0 other;
Query Match 20.5%; Score 147.4; DB 23; Length 2558;
Best Local Similarity 57.3%; Pred. No. 1.1e-27;
Matches 287; Conservative 0; Mismatches 211; Indels 3; Gaps 1;
OY 129 ACCGCGCTGCTGAGCCCACTAGTCCCGGAAACGAGTGGAGAGCCCTGCATCAGAGT 188
DB 1509 ACATGGCCAGGCGGTATGCTCTCAAGGGAGTGGGTGGCGGCGCAAT 1450
OY 189 GCGCCAGACCTGAGCTGCGCTTACGCTTATGCTGTTTCGACACGGCGGACCA 248
DB 1449 GGACCGTAGGCTGGGCAACTACCTACGCTACGCTATCACCACACCGCGGCTCCT 1390
OY 249 GCTGCAACACCCCGCGCTGCGCAGCAGAGGCGCGAATGTGACACTACACATGA 308
DB 1389 ACTGCGAGACCCGTCGCGCAGTGAACAGCCGCTGTCAGAGGCTCCAGAACTACCAATGG 1330
OY 309 AGACACTGGGCTGTGTCGACGTGGGCTACAACTTCTGATTGGAGAAAGCGGCTGTAT 368
DB 1329 ACTCCCTGGGGCTGGCGGACATCGGCTCAACTTCTGATGGCGGAGAACGGCAACGTGT 1270
OY 369 ACGAGGGCGGTGGCTGGAACCTTCAAGGGTCCCACTCAGTCACTTATGAAACCCCATGT 428
DB 1269 ACGAGGGCGGTGGCTGGAACCAATGAGCGGCCACAGCGC---CCGAGTGAACCCCTTACA 1213
OY 429 CCATTGGATACGCTTATGAGGCAACTACATGATGAGGTGGCCACACCCACAGGCATCC 488
DB 1212 GCATCGGATACGCTTCTGCGGCAACTACATGAGGACACCTTGAGGAACATGATCT 1153
OY 489 GGGAGCCAGGAGTCTACTGCGCTGCGGTGAGCTCAGAGGAGCCCTGAGTCCACTATAG 548
DB 1152 CCGCGCGCCACAGCTGTCTCAACGAGCGGCTCAACCGTGGCAGCTCAGCTCGGCTTACA 1093
OY 549 TGTCTAAGAGACCGGAGTGTGACAGCTTCTTCCAGGCAACAGCTCTTACACC 608
DB 1092 TCTGTAGGTATCGATCGCAGGTACAGCGCCAGTACCGCGGACCAATCTGTGAACG 1033

```

```

OY 609 TCATCAGATTTGGCCACACT 629
DB 1032 AGATCCGGCGCTGTCCACACT 1012
RESULT 15
ABL25876/c
ID ABL25876 standard; DNA; 4320 BP.
XX
XX AC ABL25876;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 29101.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX
XX OS Drosophila melanogaster.
XX
XX PN W0200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX DR WPI; 2001-656860/75.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX
XX PS Claim 1; SEQ ID NO 29101; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent .
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4320 BP; 1090 A; 991 C; 1012 G; 1227 T; 0 other;
Query Match 20.5%; Score 147.4; DB 23; Length 4320;
Best Local Similarity 57.3%; Pred. No. 1.2e-27;
Matches 287; Conservative 0; Mismatches 211; Indels 3; Gaps 1;
OY 129 ACCGCGCTGCTGAGCCCACTAGTCCCGGAAACGAGTGGAGAGCCCTGCATCAGAGT 188
DB 3906 ACATGGCCAGGCGCTGTATGCTCTCAAGGGAGTGGGTGGCGGCGCAAT 3847
OY 189 GCGCCAGACCTGAGCGCTGCTTACGCTATGTTGGTATGACACAGCGGCGGACAGA 248
DB 3846 GGACCGTAGGCGCTGGGCAACTACCTACGCTACGCTATCCACACACCGCGGCTCCT 3787
OY 249 GCTGCAACACCCCGCTCTGTCGCGCAGCAGAGCCCGGAATGTGACACTACACATGA 308
DB 3786 ACTGGAACACCGCTGCGCAGTGAACAGCGCGTCTCAGAGCTCCAGAACTTACACATAG 3727
OY 309 AGACACTGGGCTGTGTCGACGTGGGCTACAACTTCTGATTGGAGAAAGCGGCTGTAT 368

```

```
Db 3726 ACTCCCTGGGCTGGCCCGGACATCGGCTACACTTCTGATCGGCGGAGACGGCAACGTGT 3667
OY 369 ACGAGGGCCCTGGGCTGGGAACCTTCACGGGTGCCCACTAGGTCACTTATGGAACCCCATGT 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3666 ACGAGGGCGGCTGGGCTGGGAACATGAGGCGCCACGCCG---CCGAGTGAACCCCTACA 3610
OY 429 CCATTGGCATCAGCTTCATGGGCACTACATGATGGGTGCCACACCCAGGCCATCC 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3609 GCATGGGCTACAGCTTCCTGGGCACTACACTGGGACACCCCTGGAGCCGAACTGATCT 3550
OY 489 GGGCAGCCCGAGGCTCTACTGGGCTGTGTGGCTCAGGGAGCCCTGAGGTCCAACTATG 548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3549 CCGCGGCCAGCAGCTGCTCAACGAGAGCCGCTCAACCGTGGCCAGCTCAGCTCCGCTACA 3490
OY 549 TECTAAAGGACACCGGGATGTGCAAGCGTACACTCTCTCCAGGCAACCAAGCTTACCACC 608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3489 TCCTGTACGCTCATCGGCAAGTCAAGGGCCACCGAATGCCCGGACCCACATCTGGAAG 3430
OY 609 TCATCCAGAAATGGGCACACT 629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3429 AGATCCGGCGGTGTCCCACT 3409
```

Search completed: November 12, 2002, 00:50:55
Job time : 312 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 12, 2002, 02:29:10 ; Search time 309 seconds

(without alignments)
1392.012 Million cell updates/sec

Title: US-09-462-625-4

Perfect score: 1051

Sequence: 1 MLAMALPSLRIGAAQETE.....ISPQNQLYHLIQNMPHYRSP 191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlp
-O/cgnt2_1/USPRO.spool/US09462625/runat_07112002.101525.24830/app.query.fasta.1.327
-DB=N_Geneseq_101002 -OFAST=fastp -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Dlosum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09462625 @CGN.1.1.281 @runat_07112002.101525.24830 -NCPUT=6 -ICPUT=3
-NO_XLPHY=NO_MMAP -LARGQUERY=NEG SCORES=0 -WAIT -LONLOG=DEV-TIMEOUT=120
-MARKTIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002: *
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1051	100.0	697	21 AAC58104	Human PRO1269 uncl
2	1051	100.0	697	21 AAC37082	Human PRO1269 (UNO
3	1051	100.0	697	22 AAF54356	DNA encoding prote
4	1051	100.0	718	20 AAX21820	Human tag7 clone c
5	1051	100.0	726	21 AAA80613	Human tag7 secret
6	1051	100.0	749	21 AAA51719	Chondrosarcoma pep
7	717	68.2	688	18 AAT78509	Bovine granulocyte
8	692	65.7	549	20 AAT21819	Mouse tag7 clone c
9	648.5	61.7	677	18 AAT78510	Murine granulocyte
10	573	54.5	380	21 AAT42549	Human 5' EST isola
11	414.5	39.4	555	23 ABL25881	Drosophila melanog
12	414.5	39.4	2555	23 ABL25880	Drosophila melanog
13	413	39.3	726	24 ABA90336	Human polynucleoti
14	413	39.3	1110	22 AAD04006	Human full length
15	413	39.3	1110	24 ABA90337	Human polynucleoti
16	413	39.3	1128	23 AAD04004	Human full length
17	405.5	38.6	558	23 ABL25883	Drosophila melanog
18	405.5	38.6	639	23 ABL04835	Drosophila melanog
19	405.5	38.6	2558	23 ABL25882	Drosophila melanog
20	405.5	38.6	2639	23 ABL04834	Drosophila melanog
21	405.5	38.6	4320	23 ABL25876	Drosophila melanog
22	395.5	37.6	1876	21 AAA51718	Wound healing tiss
23	379	36.1	612	23 ABL26901	Drosophila melanog
24	371.5	35.3	457	21 AAA80663	Human secreted pro
25	348.5	33.2	1089	21 AAA15921	Human protein clon
26	348.5	33.2	1256	21 AAA15931	Human protein clon
27	348.5	33.2	1731	22 AAI67206	Nucleotide sequenc
28	348.5	33.2	1905	22 AAI67205	Nucleotide sequenc
29	348	33.1	753	19 AAV19149	Gene encoding the
30	347.5	33.1	591	19 AAV19148	Gene encoding the
31	347	33.0	1256	23 ABL14655	Drosophila melanog
32	341	32.4	1876	22 AAD02743	Murine peptidoglyc
33	339	32.3	1107	22 AAD04007	Human full length
34	339	32.3	1125	22 AAD04005	Human full length
35	338	32.2	2609	23 ABL15556	Drosophila melanog
36	338	32.2	2641	23 ABL15528	Drosophila melanog
37	331.5	31.5	9721	23 ABL07586	Drosophila melanog
38	330	31.4	561	23 ABL08767	Drosophila melanog
39	330	31.4	2561	23 ABL08766	Drosophila melanog
40	329.5	31.4	573	23 ABL15529	Drosophila melanog
41	328.5	31.3	522	19 AAV19147	Gene encoding pept
42	327	31.1	1014	23 ABL07655	Drosophila melanog
43	317	30.2	970	23 ABL25789	Drosophila melanog
44	316.5	30.1	1182	21 AAA51717	keratinocyte pepti
45	316.5	30.1	1191	21 AAT65250	Human secreted pro

ALIGNMENTS

RESULT 1

AAC58104

25-JAN-2001 (first entry)

Human PRO1269 nucleotide sequence SEQ ID NO:6.

Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;

identification; tumorigenesis; anticancer; detection; ss.

Homo sapiens.

WO200053750-A1.

14-SEP-2000.

XX 02-DEC-1999; 99WO-US28551.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WT;
XX WPI; 2000-594320/56.
DR P-PSDB; AAB24022.
XX
XX Antibodies specific for pro polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of pro
PT polypeptide activity or expression -
XX
PS Claim 50; Fig 3; 226pp; English.
XX
CC The present invention describes an antibody that binds to a human
CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO3434; PRO1927; PRO3567; PRO1293; PRO1303; PRO4354; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumors in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumors. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridizes to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human pro
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC pro polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
XX Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 5.17e-100 Length: 697
Score: 1051.00 Matches: 191
Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-462-625-4 (1-191) x AAC58104 (1-697)
QY 1 MetleuLeuAlaTPAlaLeuProSerleuLeuArgleuGlyAlaAlaGlnGlnIurhgu 20
DB 41 ATGCTGCTGGCTGGGCTCTCCAGCTCTCTTCACCTCGGAGCGGCTAGAGACAGAA 100
QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
DB 101 GACCGGCGCTGCTGACGCCCATAGTGCCTCCGAGACGAGTGAAGCCCTGCGATCAGAG 160
QY 41 CysAlaGlnHisLeuSerleuProLeuArgTrpValValSerHisThrAlaGlySer 60
DB 161 TGGCGCCAGCAGCTAGGCTGCTTACGCTATGTGTGGTATCCGACACGCGGCGACAGC 220
QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMet 80
DB 221 AGCTGCAACACCCCGCTCGTCCAGCAGCGAGCGGAGATGTGAGCAGTACACCATG 280
QY 81 LysThrLeuGlyTrpCysAspValGlyTrpAsnPhaLeuIleGlyGluAspGlyLeuVal 100
DB 281 AAGACACGCGGCTGTGAGCTGAGCTGACCTTCACTCTTATGTGAGAGAACAGCGGCTGTA 340
QY 101 TyrGluGlyArgGlyTrpAsnPhaThrGlyAlaHisSerGlyHisLeuTrpAsnProMet 120

DB 341 TACGAGGCGCGTGGCTGGAATTCACGCGGCTGCCACCTCAGGTCATCTATGGAACCCCATG 400
QY 121 SerIleGlyIleSerPheMetGlyAsnTrpMetAspArgValProThrProGlnAlaIle 140
DB 401 TCCATTGGCATGACTTCATGGGCAACTAGATGGATGGGTGCCACACCCCGGCTCATC 460
QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTrp 160
DB 461 CGGCGACGCCAGGCTCTATGGCTGCGGTGTGGCTCAGGAGCCCTGAGGTCCACTAT 520
QY 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTrpHis 180
DB 521 GTGCTCAAGGACACCGGATGTGACGCTACACTCTCTCAGGCAACACGCTCTACAC 580
QY 181 LeuIleGlnAsnTrpProHisTrpArgSerPro 191
DB 581 CTCATCCAGATTTGGCCACACTACCTCCCTCC 613
RESULT 2
AAA37082
ID AAA37082 standard; cDNA; 697 BP.
XX
AC AAA37082;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (U00639) cDNA sequence SEQ ID NO:215.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW Transmembrane; secretion; Immunoadhesion; pharmaceutical; screening;
KW ss.
XX
OS Homo sapiens.
XX
PN W0200012708-A2.
XX
PD 09-MAR-2000.
XX
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.

XX 11-JUL-1997; 97US-0893764.
PR (BOEH) BOEHRINGER INGELHEIM INT GMBH.
XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
XX WPI: 1999-120887/10.
DR P-PSDB: AAY00771.
XX
XX New nucleic acid encoding tag7 - used to inhibit tumour growth and
PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
PT leukaemia
XX
PS Claim 11: Page 126-127: 138pp: English.
XX
XX This sequence encodes the human tag7 of the invention. Cells containing
CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
CC to produce and purify antibodies; to inhibit growth of mammalian tumours,
CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
CC osteo- or fibro-sarcoma), melanoma or leukaemia; and as a molecular
CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
CC apoptosis. The tag7 coding sequences are also useful as probes for gene
CC mapping and detection of tag7 gene expression, and as primers. Antibodies
CC against tag7 are used as reagents for detecting tag7; as an antagonist of
CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
CC metastasis.
XX
XX Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 5.39e-100 Length: 718
Score: 1051.00 Matches: 191
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 20 Gaps: 0

US-09-462-625-4 (1-191) * AAX21820 (1-718)
QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGluThrGln 20
Db 68 ATGCTGCTGGCTGGGCTGCTCCAGCCCTCCTGCAGCTGGAGCGGCTAGAGACAGAA 127
QY 21 AsProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
Db 128 GACCGGCGCTGCTGCAGCCCTCAGTACTGCCCCGGAACGAGTGGAGGCCCTGCGATCAGAG 187
QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTrpValValSerHisThrAlaGlySer 60
Db 188 TGGCGCCAGCACCTAGAGCTGCTTACGCTATGTGTGTTCGCACACGGCGGCGAC 247
QY 61 SerCysAsnThrProAlaSerCysGlnGlnGlnAlaArgAsnValGlnHisThrHisMet 80
Db 248 AGCTCAACACACCCCGCTCTGTCGACGACGACGCCCGGAATGTGACAGACTACACATG 307
QY 81 LysThrLeuGlyTrpCysAspValGlyTrpAsnPhenLeuIleGlyGlnAspGlyLeuVal 100
Db 308 AAGACACTGGCGCTGGTGGACGTGGGCTACAACTTCTGATGGAGAAACGGGCTCGTA 367
QY 101 TyrGlnGlyArgGlyTrpAsnPhenThrGlyAlaHisSerGlyHisLeuThrAsnProMet 120
Db 368 TACGAGGCGCGGTGGTGAATTCACGGGTGCCACTCGAGTACTATGGAACCCCATG 427
QY 121 SerIleGlyIleSerPheMetGlyAsnTrpMetAspArgValProThrProGlnAlaIle 140
Db 428 TCCATTTGGCATAGCTTCAATGCGCAACTACATGATCGGGTCCGCCACACCGGCATC 487
QY 141 ArgAlaAlaGlnGlyLeuLeuAlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTrp 160
Db 488 CGGCGAGCCCGAGGCTCTACTGCTGGCGGTGTGGTCAAGGAGCCCTGAGGTCCAACTAT 547

QY 161 ValLeuLysGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTrpHis 180
Db 548 GTGCTCAAGAGACACCGGAGTGTGCGACTACACTCTCTCAGGCAACCAAGCTTACAC 607
QY 181 LeuIleGlnAsnTrpProHisTrpArgSerPro 191
Db 608 CTCATCCACGAATTGGCGCACATCACGCGCTCC 640

RESULT 5
AAA80613
ID AAA80613 standard; cDNA; 726 BP.
XX
XX AAA80613;
XX
XX 21-NOV-2000 (first entry)
XX
XX Human Htag7 secreted protein gene #8.
XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
XX antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
XX anticancer; vulnary; antiviral; antibacterial; antifungal;
XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
XX Crohn's disease; nephritis; hyperproliferative disorder;
XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
XX melanoma; lymphoma; wound healing; human; ss.
XX
XX Homo sapiens.
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX 27-OCT-1999; 99WO-US25031.
XX
XX 28-OCT-1998; 98US-0105971.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX NI J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
XX Greene JM;
XX
XX WPI: 2000-387742/33.
XX
XX P-PSDB: AAB25583.
XX
XX Isolated nucleic acid molecules encoding human secreted proteins are
XX used for the prevention, amelioration and treatment of autoimmune,
XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX wounds, and infectious diseases -
XX
XX Claim 1: Figure 34; 803pp: English.
XX
XX The present invention relates to 12 secreted human proteins and the
XX nucleotide sequences encoding them. The polynucleotide sequences given
XX in AAA80613 and AAB25583 encode the 12 secreted protein sequences given in
XX AAB25576-B25593. The human secreted proteins have various activities
XX dependent on the tissues in which they are expressed. Examples of the
XX activities of the proteins include: immunosuppressant;
XX anti-inflammatory; antiarthritic; antirheumatic; dermatological;
XX antiproliferative; antiarteriosclerotic; anticancer; vulnary;
XX antiviral; antibacterial; and antifungal activity. The proteins,
XX polypeptides, agonists and antagonists may be used to treat prevent
XX and/or diagnose various disease, disorders and conditions examples of
XX which include: immune disorders e.g. Addison's disease, rheumatoid
XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;
XX hyperproliferative disorders such as paraneoplasms and purpura;
XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
XX sequences may also be used in wound healing and the treatment of
XX infectious diseases. The human secreted protein gene #8 and protein
XX sequences are represented in sequences AAA80613 and AAB25583. Sequences

CC AAA80662-A80663 represent genes related to the secreted protein gene#8.
 XX Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;
 SQ Alignment Scores:
 Pred. No.: 5.47e-100 Length: 726
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-462-625-4 (1-191) x AAA80613 (1-726)

QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGluThrGlu 20
 DB 47 ATGCTCTTGGCTGGCTGCTCCAGAGCTCTTCGACTCGAGCGGCTGAGAGACGAA 106
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
 DB 107 GACCCGGCTGCTGCGAGCCCATAGTCCCGGAGAGAGTGGAGGCCCTGGCATCGAG 166
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyrValValSerHisThrAlaGlySer 60
 DB 167 TGGCCAGACAGACTGAGCCCTGCTTACGCTATGCGTATCGCACAGCGCGGCGAGC 226
 QY 61 SerCysAsnTrpProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyrHisMet 80
 DB 227 AGGTGGAACACCCCGCTGCTGCGACAGACAGCCCGGAATGGCGACACTACACATG 286
 QY 81 LysTrpLeuGlyTrpCysAspValGlyTyrAsnPhenLeuIleGlyLysAspGlyLeuVal 100
 DB 287 AAGACACTGGGCTGGTGGCGACGTGGGCTACAACTCTGATTGGAGAAAGAGGGCTGTA 346
 QY 101 TyrGluGlyArgGlyTrpAsnPhenThrGlyAlaHisSerGlyHisLeuTrpAsnPromet 120
 DB 347 TACGAGGCGCGTGGTGGAACTTCACGCGTCCACACAGTCACTATGAGAACCCCATG 406
 QY 121 SerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAlaIle 140
 DB 407 TCCATTGGCATACACTTCATAGGCACTACATGATGGGTGCCACACACCCAGGCCATC 466
 QY 141 ArgAlaIleGlnGlyLeuLeuAlaCysGlyValAlaGlnGlnGlyAlaLeuArgSerAsnTyr 160
 DB 467 CGGGACACCCAGGCTACAGGCTGCGGTGGCTCAGGAGGCCCTGAGSTCCAACTAT 526
 QY 161 ValLeuGlyGlyHisArgAspValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHis 180
 DB 527 GTGCTCAAAAGAGACACCGGATGTGACAGCTCTCTCCAGGCAACACACTCTACAC 586
 QY 181 LeuIleGlnAsnTrpProHisTyrArgSerPro 191
 DB 587 CTCATCCAGAAATGGCCACACTACCGCTCCCC 619

RESULT 6
 AAA51719 standard; cDNA; 749 BP.
 ID AAA51719;
 AC AAA51719;
 XX 31-OCY-2000 (first entry)
 DE Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
 XX
 XX Peptidoglycan recognition protein-like protein: PCR; PCR-C; regulator;
 KM chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
 KM tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
 KM inhibitor; protein co-ordinate data; ss.
 XX
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH CDS
 FT 55..645

FT FT /tag- a
 FT sig-peptide /product= PGRP-like_protein
 FT 55..117
 FT mat-peptide /tag- b
 FT 118..642
 FT /tag- c

WO200039327-A1.

XX PN 06-JUL-2000.
 XX PD
 XX PF 22-DEC-1999; 99WO-US30736.
 XX PR 23-DEC-1998; 98US-0113809.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM, Young PE, Olsen HS;
 XX DR WPI; 2000-452414/39.
 XX DR P-PSDB; AAY96964.
 XX PT Polynucleotide encoding peptidoglycan recognition protein-like protein,
 PT antibodies specific to it useful for preventing, treating conditions
 PT e.g. endotoxemic shock and auto-immune disorders and infections in mamma
 XX PS Claim 1; Fig 3; 191pp; English.

CC Novel human peptidoglycan recognition protein-like proteins (PGRP)
 CC expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
 CC or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
 CC proteins are useful for preventing, treating or ameliorating a medical
 CC condition in a mammal (claimed). PGRP is useful in augmenting the immune
 CC system in such areas as immune recognition, antigen presentation and
 CC immune system activation. Antibodies or antagonists directed against
 CC these proteins may be useful in reducing or eliminating disorders
 CC associated with tumour necrosis factor (TNF) and TNF-like cytokines,
 CC such as endotoxemic shock and autoimmune disorders and for treating
 CC infectious diseases including silicosis, sarcoidosis and idiopathic
 CC pulmonary fibrosis.
 CC XX

SQ Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;
 SQ Alignment Scores:
 Pred. No.: 5.7e-100 Length: 749
 Score: 1051.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-462-625-4 (1-191) x AAA51719 (1-749)

QY 1 MetLeuLeuAlaTrpAlaLeuProSerLeuLeuArgLeuGlyAlaAlaGlnGluThrGlu 20
 DB 70 ATGCTCTTGGCTGGCTGCTCCAGAGCTCTTCGACTCGAGCGGCTGAGAGACAA 129
 QY 21 AspProAlaCysCysSerProIleValProArgAsnGluTrpLysAlaLeuAlaSerGlu 40
 DB 130 GACCCGGCTGCTGCGAGCCCATAGTCCCGGAGAGTGGAGGCCCTGGCATCGAG 189
 QY 41 CysAlaGlnHisLeuSerLeuProLeuArgTyrValValSerHisThrAlaGlySer 60
 DB 190 TGGCCAGACAGACTGAGCTGCTTACGATGTGGTATGCGACACAGCGGCGAGC 249
 QY 61 SerCysAsnTrpProAlaSerCysGlnGlnAlaArgAsnValGlnHisTyrHisMet 80
 DB 250 AGCTGCAACACCCCGCTGCTGCGAGACAGCGCCGGAATGTVGCACACTACCAATG 309
 QY 81 LysThrLeuGlyTrpCysAspValGlyTyrAsnPhenLeuIleGlyLysAspGlyLeuVal 100
 DB 310 AAGACACTGGGCTGGTGGAGACTGCGCTACAACTCTCTGATTGGAGAAAGAGCGGCTGTA 369

CC	acids corresponding to mature BGP-A.	BGP-A and the murine
CC	homologue, MGP-A (see AAW2375), exhibit activity against	
CC	Gram-positive and Gram-negative bacteria, fungi and viruses,	
CC	specifically streptococcus aureus, Escherichia coli, Candida	
CC	albicans, Salmonella typhimurium and C. neoformans (claimed)	
CC	They can be used in human or veterinary medicine (particularly to	
CC	treat disorders associated with lipopolysaccharides, e.g. sepsis	
CC	and endotoxaemia) or as preservatives in food products or in water	
CC	supplies (claimed). They can also be applied to crops to reduce	
CC	post-harvest spoilage or expressed in transgenic plants to increase	
XX	their disease resistance. They have low immunogenicity.	
SQ	Sequence 688 BP; 126 A; 234 C; 201 G; 127 T; 0 other:	
Alignment Scores:		
Pred. NO.:	2,84e-65	Length: 688
Score:	717.00	Matches: 131
Percent Similarity:	80.75%	Conservative: 20
Best Local Similarity:	70.05%	Mismatches: 30
Query Match:	68.22%	Indels: 6
DB:	18	Gaps: 1
US-09-462-625-4 (1-191) x AAT78509 (1-688)		
OY	3 LeuAlatrrpAlalaLeuProSerIeuLeuArgLeuGlnValAalagIngturrhguAspro	22
Dd	50 CTCGGCTGGGTCnCCnCGGCCCTCCTGGCCCTCGGGCGGCTCAAGC-----	97
OY	23 AlaCyssrProllevalProrArgasnclutrrplysalaleuAlaserglucysala	42
Dd	98 -----TCCGGCACATTCGTCTCCCGGGAAAGGGGGCCCCCGCATCCAAGTGACG	151
OY	43 GlnHslSeuSerLieuProLeuArgTrValValaserHstrhalgliserSercys	62
Dd	152 CAGAGGCtAAACAACACCCTGTGCGCTACTGTGTGTGCGACACAGCGGGCACGCTGC	211
OY	63 AsnthrProAlasercySGlnGlnGlnAlaArgAnaValGlnHstryHsmetlysthr	82
Dd	212 AACACCTCGGCTCGTGCAGAGGACAGGCCCAAACACGTGCAGTACACACGTGGGAG	271
OY	83 LeuGltryPCysAspValGlyTryAsnPheLeuileGlyGluaspGlyLeuValTyrglu	102
Dd	272 CGGGCTGTGTGCGACGTGGGTACAAATTCTCTGATCGAGAAGATGGGCTCGTTAGAG	331
OY	103 GlyArngLyTPraSnPheThrGlyAlAHISsercLYHSLeuTPraSnPrometSerIle	122
Dd	332 GGCGGGGCGCTGAACACCTTAGGTGCTCATCTGTGGCCACAGTGAACCCCATRGGCATT	391
OY	123 GlyIleserPheMetGlyAsnTYMeLasPAryValProThrpProglnAlaIleargala	142
Dd	392 GGCAATTCCTTCANrGGGCACTACATCGATGGGTGCCCGCGGCTCTGCTTCAGGGCG	451
OY	143 AlAglnGlyLeuLeuAlaCYsglyValAlaGlnGlyAlaLeuArgSerAsnTyrrValleu	162
Dd	452 GCCAGAGTGTGCTGGGCTGTGTGGGCGAGCTCGGGGAATCACTGACTCACTAGCAAGTC	511
OY	163 LysGLyHisArqspValGlnArgThrlenuSerProGlysnGlnleutyRHLSleuIle	182
Dd	512 AAAGAGCACCGCGATGTGACGACCAACGCTCTCTCCAGGGGACGAGCTGTAAAATATC	571
OY	183 glnAsnTrpProHistryArg 189	
Dd	572 CAGCAGTGGCGGCACCTACCGG 592	
RESULT 8		
ID	AAX21819 standard; cDNA; 549 BP.	
XX	AAX21819;	
XX	18-MAY-1999 (first entry)	
DE	Mouse tag7 clone coding sequence.	

CC	acids corresponding to mature BGP-A.	BGP-A and the murine
CC	homologue, MGP-A (see AAW2375), exhibit activity against	
CC	Gram-positive and Gram-negative bacteria, fungi and viruses,	
CC	specifically streptococcus aureus, Escherichia coli, Candida	
CC	albicans, Salmonella typhimurium and C. neoformans (claimed)	
CC	They can be used in human or veterinary medicine (particularly to	
CC	treat disorders associated with lipopolysaccharides, e.g. sepsis	
CC	and endotoxaemia) or as preservatives in food products or in water	
CC	supplies (claimed). They can also be applied to crops to reduce	
CC	post-harvest spoilage or expressed in transgenic plants to increase	
XX	their disease resistance. They have low immunogenicity.	
SQ	Sequence 688 BP; 126 A; 234 C; 201 G; 127 T; 0 other:	
Alignment Scores:		
Pred. NO.:	2,84e-65	Length: 688
Score:	717.00	Matches: 131
Percent Similarity:	80.75%	Conservative: 20
Best Local Similarity:	70.05%	Mismatches: 30
Query Match:	68.22%	Indels: 6
DB:	18	Gaps: 1
US-09-462-625-4 (1-191) x AAT78509 (1-688)		
OY	3 LeuAlatrrpAlalaLeuProSerIeuLeuAryLeuGLyAlaLaInglutrhngluAspro	22
Db	50 CTCGGCTGGGTCnCCnCGCCCTCCTGGCCCTCGGGCGGCTCAAGC-----	97
OY	23 AlaCyssrProllevalPrOArAsnclutrpLyAlaleuAlaserglucysala	42
Db	98 -----TCCGGACATTCGTGCCGGGAAGGGGGCCCCCGCATCCAAGTGcAGC	151
OY	43 GlmHslSeSerLieuProLeuArgTrValVaIalSerHsthrIaglyserSercys	62
Db	152 CAGAGGCtAAACAACACCCTGTGCGCTACTGTGTGTGCGCACACGGCGGACAGCTGC	211
OY	63 AsnthrProAlasercysgLnGLInGlAlaArghAnaVlglnHstYrHsmetLysthr	82
Db	212 AACACCTCGGCTCGTGCAGAGGACAGGCCCAAACACSTGCAGTACACACGTGGGGAG	271
OY	83 LeuGLITPCysAspValGLyTYrAsnPheLeuileGLyluspglyLeuValTYrglu	102
Db	272 CGGGCTGGTGCAGACGTGGGTACAAATTCTCTGATCGAGAAGATGGGCTCGTATGAG	331
OY	103 GLYArngLYTPRAsnpheThrGLYAlAHISserLIHSLeuTrPaSPrometSerIle	122
Db	332 GGCGGGGCGCTGAAACACCTTAGGTGCTCATCTGTGGCCACAGTGAACCCCATRGCCATC	391
OY	123 GlyIleserPheMetGLYAsnTYMeLasPARyVALProThrPRoGlnAlaIleargala	142
Db	392 GGCAATTCCTTCANtgGGGAActAcATGCAnTGSGGTGCCCGCGGCTCTGCTTCAGGGCG	451
OY	143 AlAgInglyLeuLeuAlaCYsgLYValAlaGLInGLYAlaLeuArGSerAsnyrValLeu	162
Db	452 GCCAGAGTCTGCTGGGCTGTGGGGCAGCTCGGGGAATACCTGACTCTACCTACGAAATC	511
OY	163 LysGLYHsArqspValGLInARgThrLeuSerProLIYsnGLInleuTYrHLSleuIle	182
Db	512 AAAGAGCACCGCGATgtGACGACAAACSCCTCTTCACGGGACGAGCTGTATAAAATCATC	571
OY	183 glnAsnTrPrOHistTyArg 189	
Db	572 CAGCAGTGGCGGCACTACCGG 592	
RESULT 8		
ID	AAX21819 standard; cDNA; 549 BP.	
XX	AAX21819;	
XX	18-MAY-1999 (first entry)	
DE	Mouse tag7 clone coding sequence.	

QY	61	Se	Cys	Asn	Thr	Pro	Ala	Ser	Cys	Gln	Gln	Gln	I	Ala	Arg	Asn	Val	Gln	I	His	Tyr	His	Met	80						
Db	228	AG	CTG	CA	AC	CA	CC	CC	CG	CT	CG	TG	TG	CC	AG	CA	AG	GG	CC	GG	AA	TG	TG	CA	CT	AC	CA	TG	287	
QY	81	Lys	Thr	Leu	Gly	Tyr	Cys	Asp	Val	Gly	Tyr	Asn	Phe	Leu	Ile	Gly	Glu	Asp	Gly	Leu	Val	100								
Db	288	AAG	ACA	CTG	GGG	CTG	TG	TG	CG	AC	CT	G	CT	CA	ACT	KCC	TG	ATT	G	CA	GAG	AC	GCG	GCT	CGT	347				
QY	100	1	T	G	T	G	C	T	G	A	T	G	A	T	G	A	T	G	A	T	G	A	T	G	A	T	G	A	T	G
Db	348	AT	AC	AA	G	GG	CC	CG	CT	G	GT	G	CA	CT	CA	C	CG	CG	G	T	378									
RESULT 11																														
ID	ABL25881	standard; DNA; 555 BP.																												
XX	ABL25881;																													
XX	26-MAR-2002	(first entry)																												
DE	Drosophila melanogaster	genomic polynucleotide SEQ ID NO 29116.																												
KW	Drosophila; developmental biology; cell signalling; insecticide;																													
XX	pharmaceutical; gene; ds.																													
OS	Drosophila melanogaster.																													
PN	WO200171042-A2.																													
XX	27-SEP-2001.																													
PE	23-MAR-2001; 2001WO-US09231.																													
XX	23-MAR-2000; 2000US-191637P.																													
PR	11-JUL-2000; 2000US-0614150.																													
XX	(PEKE) PE CORP NY.																													
PA	Venter JC, Adams M, Li PWD, Myers EW;																													
P1	WPI; 2001-656860/75.																													
DR	New isolated nucleic acid detection reagent for detecting 1000 or more																													
PT	genes from Drosophila and for elucidating cell signalling and cell-cell																													
PT	interactions -																													
XX	Claim 1; SEQ ID NO 29116; 21pp + Sequence Listing; English.																													
PS	The invention relates to an isolated nucleic acid detection reagent																													
CC	capable of detecting 1000 or more genes from Drosophila. The invention is																													
CC	useful in developmental biology and in elucidating cell signalling and																													
CC	cell-cell interactions in higher eukaryotes for the development of																													
CC	cell-cell interactions in higher eukaryotes for the development of																													
CC	disclases genomic DNA sequences (AB16176-AB130511), expressed DNA																													
CC	sequences (AB101840-AB16175) and the encoded proteins																													
CC	(AB57737-AB872072).																													
CC	The sequence data for this patent did not form part of the printed																													
CC	specification, but was obtained in electronic format directly from WIPO																													
CC	at ftp.wipo.int/pub/published_pct_sequences.																													
XX	Sequence 555 BP; 120 A; 192 C; 140 G; 103 T; 0 other;																													
Alignment Scores:																														
Pred. No.:	6.26e-34	Length:	555																											
Score:	414.50	Matches:	77																											
Percent Similarity:	66.26%	Conservative:	31																											
Best Local Similarity:	47.24%	Mismatches:	54																											
Query Match:	39.44%	Indels:	1																											
DB:	23	Gaps:	1																											
US-09-462-625-4 (1-191) x ABL25881 (1-555)																														

Db	67	ATCATCTCCAAATCGGAGGAGGGGCGCGCTTCCGCCACAGCAAGACCTCGCTGGCCAAAC	126
Qy	48	ProteinArgTyrValValValSerHisThrAlaGlySerSerCysAsnThrProAlaSer	67
Db	127	TACCTGAGCTACGCGGTGATCCACACACACCGGTGAAACTACTGACAGCAACGAGCGGCC	186
Qy	68	CysGlnGlnGlnAlaArgAsnValAlaGlnHisTyrHisMetIleThrLeuGlyTyrPcysAsp	87
Db	187	TGCATCCACACACCTGCAGAACATCCAGAGCGCTCCACATGTGACTCTCCGTGGAGTGGCCGAT	246
Qy	88	ValGlyTyrAsnPheLeuLeuIleGlyIuAspGlyLeuValTyrGluGlyTyrGlyTyrAsn	107
Db	247	ATCGGCTCAACTTCTCTGATCCGCGGAGACACGGCAGACGTTACGAGGCTGCGGCTGGAAC	306
Qy	108	PheThrGlyAlaHisSerGlyHisIleuTyrAsnProMetSerIleGlyIleSerPheMet	127
Db	307	GTTATGGGTGCTCACGCCCACTAAC--TGAACCTCCAAAGTCTATCGGACATCTCTTCCTG	363
Qy	128	GlyAsnTyrMetLeuAspArgValProThrPheProGlnAlaIleArgAlaAlaGlnGlyLeu	147
Db	364	GCCACATCAATATACCAACACCCCTCACCTCTGCATACCGGCTGCCAAGGCTGTGCTC	423
Qy	148	AlaCysGlyValAlaGlnGlnIleValLeuArgSerAsnTyrValLeuIleGlnAsnTyrProHis	167
Db	424	TCCGATGCGGTGATGCGCGGCCAGATCGTTTCCGGATATACCTCTGTACGACATCGGCAG	483
Qy	168	ValGlnArgTyrThrLeuSerProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTyrProHis	187
Db	484	GTCGGCTCCACCGAGTCCCGGCCGACCAACATCTGGAAGACAGATCCGACCTGTGTCACAC	543
Qy	188	TyrArgSer 190	
Db	544	TGGAAAGGCC 552	
RESULT 12			
ID	ABL25880/C		
	ABL25880 standard; DNA; 2555 BP.		
XX	ABL25880;		
XX	AC		
XX	DT		
XX	26-MAR-2002 (first entry)		
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 29113.		
XX	XX		
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ds.		
XX	XX		
OS	Drosophila melanogaster.		
XX	XX		
PN	WO200101042-A2.		
PD	27-SEP-2001.		
XX	XX		
PF	23-MAR-2001; 2001WO-US09231.		
XX	XX		
PR	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX	XX		
PA	(PEKE) PE CORP NY.		
XX	XX		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
XX	XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
XX	interactions -		
XX	XX		
PS	Claim 1; SEQ ID NO 29113; 21pp + Sequence Listing; English.		
CC	XX		
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		

CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB1016-ABJ30511), expressed DNA
 CC sequences (AB101840-ABL6175) and the encoded proteins
 CC (AB101737-AB101702).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 2555 BP: 708 A; 470 C; 559 G; 818 T; 0 other;

Alignment Scores:

Pred. No.:	4,91e-33	Length:	2555
Score:	414.50	Matches:	77
Percent Similarity:	66.268	Conservative:	31
Best Local Similarity:	47.248	Mismatches:	54
Query Match:	39.448	Indels:	1
DB:	23	Gaps:	1

US-09-462-625-4 (1-191) x ABL25880 (1-2555)

QY 28 TLevalProArhAnGluTrpLysAlaLeuAlaSerGlnCysAlaGlnHisLeuSerLeu 47
 DB 1489 ATCATCTCCAGTCGAGTGGGGCCCTCCGCCACAGACAGAAAGCTCGGGCCAC 1430
 QY 48 ProLeuArgTyValValAlaSerHisThrAlaGlySerGlySerGlySerGlySer 67
 DB 1429 TACCTTAGCTAGCCCGTGTATCCACACCGCGTGAACATGACGACACCAAGCGCC 1370
 QY 68 CysGlnGlnGlnAlaArgAsnValGlnHisTyrHisMetLysThrLeuGlyTrpCysAsp 87
 DB 1369 TGCATACACAGGTGAGAAATCATCCAGGCTACACATGATCCTCGGGCTGGCCGAT 1310
 QY 88 ValGlyTyrAsnPhelLeuIleGlyIleuAspGlyLeuValTyrGluGlyArgGlyTrpAsn 107
 DB 1309 ATCGGCTACAACTTCCTGATCGCGGAGACGCAAGCTGACAGAGGTGCGGCTGGAAC 1250
 QY 108 PheThrGlyAlaHisSerGlyHisLeuTrpAsnPrometSerIleGlyIleSerPheMet 127
 DB 1249 GTTAGGGGTCTCACCCCACTAAC--TGGAACTCCAAAGTCTGCGCACTCTCTCTG 1193
 QY 128 GlyAsnTyrMetAspArgValProThrProGlnAlaIleArgAlaIleGlnGlyLeuLeu 147
 DB 1192 GGCAGACTACAAATACCAACACCCCTGCTGCTGCAATCCCTGCGCAAGGCTGCTC 1133
 QY 148 AlaCysGlyValAlaGlnGlyAlaLeuArgSerAsnTyrValLeuYsglyHisArgAsp 167
 DB 1132 TCCGATGCGCGTACGTCGCGCAGATCGTTTCCGATACATCTGTACGACATCGGCAG 1073
 QY 168 ValGlnArgThrLeuSerProGlyAsnGlnLeuTyrHisLeuIleGlnAsnTrpProHis 187
 DB 1072 GTGCGCTCCACCGAGTGCCTCCGACCAACATCTGGAACGATCGCACCTGTGTCAC 1013
 QY 188 TyrArgSer 190
 DB 1012 TGGAGGCC 1004

RESULT 13

ABA90336 standard; cDNA; 726 BP.

XX ABA90336;
 XX 12-FEB-2002 (first entry)
 XX Human polynucleotide #11.

KW Human; neotropic; neuroprotective; anticonvulsant; antidepressant;
 KW neuroleptic; tranquiliser; antiarrhythmic; cardiacy; antisthmatic;
 KW antiinflammatory; antidiabetic; hepatotropic; virucide; antidiabetic;
 KW nephrotropic; anorectic; cyostatic; vaccine; neurological disease;
 KW cardiovascular disease; respiratory disease; liver disease;

KW renal disease; skeletal muscle disease; gastrointestinal disease;
 KW placental disease; testicular cancer; male fertility; pancreatic disease;
 KW sa.

XX Homo sapiens.

XX WO200181363-A1.

XX 01-NOV-2001.

XX 26-APR-2001; 2001WO-US13360.

XX 27-APR-2000; 2000US-199963P.

XX 11-MAY-2000; 2000US-203336P.

XX 25-MAY-2000; 2000US-207087P.

XX 26-MAY-2000; 2000US-207546P.

XX (SMIR) SMITHKLINE BEECHAM CORP.

XX (SMIR) SMITHKLINE BEECHAM PLC.

XX Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kohnick KS;
 XX Lai Y, Xie Q;
 XX WPI; 2002-041392/05.

XX P-PSDB; ABB53271.

XX Novel polypeptides and polynucleotides useful as a vaccine for
 XX preventing and treating diseases associated the polypeptide, e.g.
 XX Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,
 XX asthma, amnesia

XX Claim 2; Page 51; 116pp; English.

CC The invention relates to an isolated polypeptide comprising a 277, 480,
 CC 583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,
 CC 844, 782, 262, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,
 CC 784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as
 CC given in the specification. The polypeptides, modulators of the
 CC polypeptides and antibodies against the polypeptides are useful for
 CC treating diseases such as neurological and psychiatric diseases
 CC including Alzheimer's, parasympathetic palsy, Huntington's disease,
 CC myotonic dystrophy, anorexia and depression; cardiovascular diseases,
 CC including congestive heart failure, Hodgkin's disease and myocardial
 CC infarction; respiratory diseases including asthma, chronic obstructive
 CC pulmonary disease, cystic fibrosis and adult respiratory distress
 CC syndrome; liver diseases including hypercholesterolemia, cirrhosis,
 CC viral and nonviral hepatitis, Type II diabetes mellitus, and impaired
 CC glucose tolerance; renal disease including renal failure, acute tubular
 CC necrosis and glomerulonephritis; skeletal muscle diseases including
 CC Bardenburg's disease, hypoglycaemia and obesity; gastrointestinal
 CC diseases including myotonia congenita and intestinal obstruction; lymph
 CC diseases including lymphoglandularia; diseases of placenta including
 CC chorioncarcinoma; diseases of testes including testicular cancer,
 CC male reproductive diseases including low testosterone and male
 CC infertility; and disease of pancreas including diabetic ketoacidosis,
 CC Type 1 and 2 diabetes and obesity. The present sequence encodes a
 CC polypeptide of the invention.

CC Sequence 726 BP: 152 A; 215 C; 198 G; 161 T; 0 other;

Alignment Scores:

Pred. No.:	1,29e-33	Length:	726
Score:	413.00	Matches:	76
Percent Similarity:	62.11%	Conservative:	42
Best Local Similarity:	40.00%	Mismatches:	66
Query Match:	39.30%	Indels:	6
DB:	24	Gaps:	3

US-09-462-625-4 (1-191) x ABA90336 (1-726)

QY 4 AlatrPlalaleProSerLeuLeuArg-----LeuGlyAlaAlaGlnGluThr 19
 DB 157 AGTATCTTAGCCACTCTTGTGAAAGCGGAGAACTGCTGCGCCCTCGGACAGACAA 216

Db	838	ATTATTTGAAGGGTGGCGTGGATGTTCACAGCTCC---TCCACCCCTGGCTACGATGAC	894
Oy	120	MetSerIleGlyIleSerPheMetGlyAsnTyrMetAspArgValProThrProGlnAla	139
Db	895	ATTGCCCTGGCATTTACCTTCATCGGCACCTTCACAGAGTATACACCCAACTGCTGACGA	954
Oy	140	IleAArgAlaIaIaGlnGlyLeuLeuAlaCysGlyValAlaIaGlnGlyAlaLeuAArgSerAsn	159
Db	955	CTTAGAGGCCACCCCAAGACCTGATCCAGTGTGCCATGTCACAGGTACCTACCTCCACAC	1014
Oy	160	TyrValIleuLeuGlyHisAArgAspValGlnArgThrIleuSerProGlyAsnGlnLeuTyr	179
Db	1015	TACCTGCTGTGGTGGCCACAGTATGTGTGGCCCAACCTGTCTCTGGGACAGCTTTATAC	1074
Oy	180	HisLeuIleGlnAsnTyrProHisTyrArg	189
Db	1075	AACATCATCAGCACCGCTGGCTCATTTCCAA	1104
RESULT 15			
ABA90337			
ID	ABA90337	standard; CDNA; 1110 BP.	
XX	AC	ABA90337;	
XX	AB		
DT	12-FEB-2002	(first entry)	
XX	DE		
XX	DE	Human polynucleotide #12.	
KW	Human; nocotropic; neuroprotective; anticonvulsant; antidepressant;		
KW	neuroleptic; tranquilliser; antiarrhythmic; cardiant; antidiabetic;		
KW	antiinflammatory; antilipemic; hepatotropic; virucide; antidiabetic;		
KW	neurotropic; anorectic; cyostatic; vaccine; neurological disease;		
KW	cardiovascular disease; respiratory disease; liver disease;		
KW	renal disease; skeletal muscle disease; gastrointestinal disease;		
KW	placental disease; testicular cancer; male fertility; pancreatic disease;		
XX	ss.		
XX	OS	Homo sapiens.	
PN	XX	WO200181363-A1.	
XX	PD	01-NOV-2001.	
XX	PD	26-APR-2001; 2001WO-US13360.	
XX	PR	27-APR-2000; 2000US-199963P.	
PR	PR	11-MAY-2000; 2000US-203336P.	
PR	PR	25-MAY-2000; 2000US-207087P.	
XX	PR	26-MAY-2000; 2000US-207546P.	
PA	PA	(SMIK) SMITHKLINE BECHAM CORP.	
PA	PA	(SMIK) SMITHKLINE BECHAM PLC.	
XX	XX		
XX	XX	Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z, Kabnick KS;	
PI	PI	Lai Y, Xie Q;	
XX	XX		
DR	DR	WPI; 2002-041392/05.	
XX	XX	P-PSDB; ABB53272.	
PT	PT	Novel polypeptides and polynucleotides useful as a vaccine for	
PT	PT	preventing and treating diseases associated the polypeptide, e.g.	
PT	PT	Alzheimer's disease, dyslipidemia, obesity, diabetes, infertility,	
PT	PT	asthma, amnesia -	
XX	XX		
PS	PS	Claim 2; Page 51; 116pp; English.	
CC	CC	The invention relates to an isolated polypeptide comprising a 277, 480,	
CC	CC	583, 581, 628, 424, 638, 229, 310, 841, 241, 369, 382, 185, 586, 1026,	
CC	CC	844, 782, 628, 394, 471, 485, 286, 533, 495, 350, 619, 490, 462, 255,	
CC	CC	784, 252, 593, 472, 607, 781, 640, 686 or 154 amino acid sequence as	
CC	CC	given in the specification. The polypeptides, modulators of the	
CC	CC	polypeptides and antibodies against the polypeptides are useful for	

treating diseases such as neurological and psychiatric diseases
including Alzheimer's, paraneuronal palsy, Huntington's disease,
myotonic dystrophy, anorexia and depression; cardiovascular diseases
including congestive heart failure, Hodgkin's disease and myocardial
infarction; respiratory diseases including asthma, chronic obstructive
pulmonary disease, cystic fibrosis and adult respiratory distress
syndrome; liver diseases including hypercholesterolemia, cirrhosis,
viral and nonviral hepatitis, type II diabetes mellitus, and impaired
glucose tolerance; renal disease including renal failure, acute tubular
necrosis and glomerulonephritis; skeletal muscle diseases including
Eulenburg's disease, hypoglycemia and obesity; gastrointestinal
diseases including lymphangiectasia; diseases of placenta including
choriochorionoma; diseases of testes including testicular cancer,
male reproductive diseases including low testosterone and male
infertility, and disease of pancreas including diabetic ketacidosis,
type I and 2 diabetes and obesity. The present sequence encodes a
polypeptide of the invention.

SQ Sequence 1110 BP; 249 A; 319 C; 295 G; 247 T; 0 other;

Alignment Scores:	
Pred. No.:	2, 28e+33
Score:	413.00
Percent Similarity:	62.18
Best Local Similarity:	40.008
Query Match:	39.30%
DB:	29
	Gaps: 3
	Length: 1110
	Matches: 76
	Conservative: 42
	Mismatches: 66
	Indels: 6
	Gaps: 3

US-09-462-625-4 (1-191) x ABA90337 (1-1110)

[illegible]

Search completed: November 12, 2002, 04:15:43

Wed Nov 13 09:37:10 2002

us-09-462-625-4.rng

Page 14

Job time : 317 secs
